
NAVAL MEDICAL RESEARCH UNIT DAYTON

# *Predicting Performance during Chronic Sleep Loss: Identification of Factors Sensitive to Individual Fatigue Resistance*

BETH M. HARTZLER, PH.D., JOSEPH F. CHANDLER, PH.D.,
CHELSEA B. S. LEVIN, B. A., & ASHLEY E. TURNMIRE, M.S.

NAMRU-D REPORT NUMBER 15-47

Reviewed and Approved
18 MARCH 2015

Jeffrey M. Andrews, CAPT, MSC, USN
Commanding Officer

*The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.*

*This work was funded by work unit number 71003*

*The study protocol was approved by the Naval Medical Research Unit Dayton Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects. Project Number: NAMRL.2010.0001.*

# REPORT DOCUMENTATION PAGE



**1. REPORT DATE (DD MM YY)**
18 MAR 2015

**2. REPORT TYPE**
Technical Report

**3. DATES COVERED (from – to)**
JAN 2011 – MAR 2015

**4. TITLE**
Predicting performance during chronic sleep loss: Identification of factors sensitive to individual fatigue resistance

**5a. Contract Number**:
**5b. Grant Number**:
**5c. Program Element Number**:
**5d. Project Number:**
**5e. Task Number:**
**5f. Work Unit Number:71003**

**6. AUTHORS**
Beth M. Hartzler[1], Joseph F. Chandler[2], Chelsea B. S. Levin[1], and Ashley E. Turnmire[3]

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
[1]Naval Medical Research Unit Dayton, 2624 Q St., Bldg 851, Area B, Wright-Patterson AFB, OH 45433-7955
[2]Birmingham-Southern College, 900 Arkadelphia Rd., Birmingham, AL 35254
[3]Naval Medical Research Unit San Antonio, 3650 Chambers Pass, JBSA Fort Sam Houston, TX 78234

**8. PERFORMING ORGANIZATION REPORT NUMBER (PAO to assign No.)**
Report No. NAMRU-D-15-47

**8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)**
Office of Naval Research
One Liberty Center
875 N Randolph St., Ste. 1425
Arlington, VA 22203-1995

**10. SPONSOR/MONITOR'S ACRONYM(S)**
ONR

**11. SPONSOR/MONITOR'S REPORT NUMBER(s)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.
I am an employee of the U.S. Government. This work was prepared as part of my official duties. Title 17 U.S.C. §105 provides that 'Copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

**14. ABSTRACT**

Fatigue due to sleep loss has long been recognized as an insidious threat to military performance. In an effort to mitigate the detrimental effects of fatigue on Warfighter performance, generalized biomathematical models have been developed to estimate fatigue-related performance impairments for a given schedule. However, these models fail to account for individual differences in fatigue susceptibility. To address this shortcoming, the study described herein utilized a chronic sleep restriction paradigm (4 h time-in-bed for 4 nights) with the goal of identifying neurobehavioral measures sensitive to the effects of sleep restriction, and which thus might be beneficial in developing a predictive algorithm sensitive to individual differences. Group-level analyses revealed a number of measures which were sensitive to increasing fatigue across the duration of the study, such as the Psychomotor Vigilance Task (PVT), Profile of Mood States (POMS), and certain oculometric patterns. Individual-level analyses further revealed that several of these factors were also sensitive to differences in fatigue susceptibility. Moreover, the predictive value of the Fatigue Avoidance Scheduling Tool (FAST) was increased ten-fold by combining the performance estimates other assessments of the individual's present state. Successful improvement and subsequent use of these types of algorithms could help optimize both mission efficacy and safety by identifying personnel who are best able to maintain performance under fatigued operational conditions.

**15. SUBJECT TERMS**
Fatigue; sleep loss; performance; predictive models

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 18a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | CAPT Jeffrey M. Andrews<br>Commanding Officer<br>Naval Medical Research Unit Dayton |
| UNCL | UNCL | UNCL | UNCL | 51 | 18b. TELEPHONE NUMBER (INCLUDING AREA CODE)<br>COMM/DSN: 937-938-3872 (DSN: 798) |

**Standard Form 298 (Rev. 8-98)**
***Prescribed by ANSI Std. Z39-18***

# Predicting Performance during Chronic Sleep Loss: Identification of Factors Sensitive to Individual Fatigue Resistance

Beth M. Hartzler, Ph.D., Joseph F. Chandler, Ph.D., Chelsea B. S. Levin, B. A.,
& Ashley E. Turnmire, M.S.

*The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.*



*The study protocol was approved by the Naval Medical Research Lab Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects. Project Number: NAMRL.2010.0001.*

*This work was supported by Work Unit Number 71003.*

Acknowledgements

The authors would like to thank Dr. J. Lynn Caldwell and Mrs. M. Bernadette VerStraten for their helpful comments on this report.

## EXECUTIVE SUMMARY

### Background

Fatigue due to sleep loss is a significant problem for the modern workforce (Rajaratnam & Arendt, 2001) and is the most frequently-cited contributing factor in Naval Aviation mishaps (Belland, 2012; Naval Safety Center, 2014). Addressing well-founded concerns about insufficient sleep among military personnel, analeptics have been made available within each of the service branches, and guidelines have been developed regarding ideal sleeping conditions (Caldwell & Caldwell, 2005; Gore, Webb, & Hermes, 2010). Regrettably, fatigue remains a problem, making it clear that better tools are necessary for fatigue prevention and detection. In response to this need, computer models have been developed which can predict appropriate work/sleep cycles to minimize fatigue (e.g., Sleep, Activity Fatigue, and Task Effectiveness [SAFTE™], which is the basis for both FlyAwake© and Fatigue Avoidance Scheduling Tool, FAST™), yet these models fail to take into account important individual differences in fatigue states and susceptibility to fatigue. However, research has determined that there are significant individual differences with regard to fatigue (Killgore, Grugle, Reichardt, Killgore, & Balkin, 2009; Van Dongen, Baynard, Maislin, & Dinges, 2004; Van Dongen, Caldwell, & Caldwell, 2006), suggesting a strong need for an individualized assessment of readiness-to-fly.

### Purpose

Research using an acute sleep loss scenario has established that the PMI FIT 2000 (PMI) as well as several measures from the Flight Fit battery can be effective tools for detecting fatigue due to sleep loss. However, these tools need to be validated to ensure their accuracy for chronic sleep loss as well. The present study was designed to address this need by reducing participants' nightly time in bed to 4 hours for 4 consecutive days while evaluating performance on a variety of different tasks. The PMI is an instrument designed to gauge physiologic state by measuring occulometric characteristics, such as saccadic velocity and pupil diameter. The Flight Fit is a brief neuropsychological battery which evaluates abilities such as short-term memory and visual scanning. Additionally, to determine which measures would be correlated with performance on the Psychomotor Vigilance Task (PVT), widely used to assess vigilance and attention, participants were tested on a flight simulator, X-Plane; voice analysis, variations in which can be used to monitor fatigue; the dual *n*-back task, which measures executive function; the Revised NEO Personality Inventory, which assesses personality traits such as neuroticism and extraversion; the Stanford Sleepiness Scale, a subjective assessment of fatigue; the Profile of Mood States (POMS), a tool to evaluate fluctuations in active mood state; and the University of Pennsylvania Smell Identification Test because previous work has found a relationship between extended time awake and a decrement in the ability to identify odors.

### Method

The performance of 24 participants on the above tasks was observed over the course of 8 days. Participants were trained on each of the tasks during the first day of the study, baseline measures were taken during the second day of the study, and then 2 days later participants returned to the lab and remained there for 4 consecutive days. During these 4 days, participants were limited to 4 hours of time in bed per 24 hours, and their performance in response to fatigue was evaluated five times per day. It was hypothesized that participants' performance on the PMI, an oculometric evaluation, and the Flight Fit, a cognitive performance measure, could be used to predict an individual's susceptibility to fatigue, based on performance on the PVT.

Additionally, it was further hypothesized that other measurements such as performance on a flight simulator, voice analysis, executive function, mood, and odor identification, would correlate with performance on the PVT over the course of the study.

**Results**

Results of the initial repeated measures analyses of variance indicated that several measures (PMI Constriction Amplitude and total lapse time on the X-Plane flight simulator) were highly sensitive to troughs in the circadian cycle. Conversely, PMI Pupil Diameter appeared to be more influenced by homeostatic drive, which is the pressure to sleep that gradually increases with continuous wakefulness. Finally, there were a number of measures which were able to effectively track both the circadian and homeostatic processes over the period of sleep restriction including the Stanford Sleepiness Scale (SSS), components of the POMS and PVT, and PMI Saccadic Velocity.

Subsequent analyses developed hierarchical linear models to identify which factors were able to predict fatigue at both the group and individual level, as measured using the lapse measure from the PVT. Significance at the group level confirmed that changes in response to sleep loss were evident, whereas significance at the individual level revealed inter-individual variability in response to sleep restriction. Results indicated that several factors were only significant at the group level (POMS Vigor-Activity and SSS), while others were only significant at the individual level (PMI Pupil Diameter, Constriction Latency, and Constriction Amplitude). The FAST performance estimates and flight simulator total lapse time, as well as POMS Fatigue/Inertia and Total Mood Disturbance, and PMI Saccadic Velocity, were significant at both group and individual levels in predicting PVT lapses over the course of the study.

These five factors which were significant at both levels were further examined through a series of enter-method linear regression analyses to determine which combination might best predict changes in fatigue as measured by the number of PVT lapses. In all of these analyses, FAST performance estimates alone predicted very little of the variance, though the algorithm was strengthened by the four other factors listed above. Moreover, grouping the participants based on their rank (i.e., top 25%, middle 50%, or bottom 25%) for the personality facets Gregariousness and Activity led to an even better fit of the predictive algorithm. Similar results were obtained by grouping the data based on time of test administration (0730, 1530, and 2330), with the strongest algorithm using data from the early morning test session.

**Discussion**

Taken together, these findings suggest that basic subjective, cognitive, and physiologic tasks are sensitive to changes in behavior and performance as fatigue increases, and the combination is essential both for optimal assessment and prediction of impairments due to sleep loss. Additionally, these results supported previous work which concluded that, although there are some similarities, changes in response to chronic sleep restriction can be quite different from what is observed during total sleep deprivation.

## INTRODUCTION

Military aviators encounter numerous dangers but one of the most prevalent threats is the effect of fatigue due to sleep loss. Fatigue has been recognized as the number one aeromedical factor implicated in Naval Aviation flight mishaps (Naval Safety Center, 2014). In addition to the commonly experienced impairments such as a decreased ability to focus attention, solve problems, and remember instructions, as well as increased stress, there are some decrements which may be of particular concern to pilots (Hartzler, 2014). Specifically, pilots experiencing sleep loss exhibit visual neglect for both the peripheral and central fields (Kendall et al., 2006; Rogé et al., 2003) as well as increased risk-taking behavior (Killgore et al., 2006; Venkatraman et al., 2007) and confusion (Drury et al., 2012). However, sustained or continuous operations in high tempo, wartime operations often result in significant sleep loss, consequently making fatigue inevitable.

A large body of research has concluded that although everyone is susceptible to fatigue as a result of sleep loss, the degree to which fatigue hinders performance varies widely among individuals (e.g., Bliese, Wesensten, & Balkin, 2006; Killgore, Grugle, Reichardt, Killgore, & Balkin, 2009; Rupp, Wesensten, Bliese, & Balkin, 2009; Van Dongen, Baynard, Maislin, & Dinges, 2004; Van Dongen, Vitellaro, & Dinges, 2005), and that these differences are stable, substantial, and independent of each individual's recent sleep history (Van Dongen & Belenky, 2009). Further, authors have concluded that sleep restriction may be used to reveal significant inter-individual differences in neurobehavioral functioning (Banks & Dinges, 2007), but caution that it should not be assumed that a fatigued participant who demonstrates impaired performance on one task will have a comparable impairment on all other tasks (Van Dongen, Baynard, Maislin, & Dinges, 2004).

The inter-individual differences in fatigue susceptibility range from physical and observable traits to neurobehavioral characteristics. For example, several studies have found that in response to chronic sleep restriction, young adults demonstrate poorer cognitive performance than do older adults, yet younger adults return to baseline quality performance more quickly (Bliese, Wesensten, & Balkin, 2006; Rupp, Wesensten, Bliese, & Balkin, 2009; Sato, Kawada, Ogawa, Aoki, & Suzuki, 1993). Similar differences have also been found based on occupation (Caldwell et al., 2005) and personality traits (Killgore, Richards, Killgore, Kamimori, & Balkin, 2007). Inter-individual differences in susceptibility have also been noted for cortical arousal, with the performance of those demonstrating higher levels of both global arousal (Caldwell et al., 2005; Killgore et al., 2007) and prefrontal activation (Killgore et al., 2009) typically being less impaired as a result of sleep loss than are participants demonstrating lower levels of activation. Finally, chronotype differences (i.e., morningness versus eveningness) have also been shown to influence susceptibility to fatigue, depending on the circadian phase position during which the individual's performance is evaluated (Van Dongen & Dinges, 2005).

Because individuals have proven to be poor judges of their own level of fatigue (Banks & Dinges, 2011; Van Dongen, Baynard, Maislin, & Dinges, 2004), tools have been developed which are designed to aid in determining when a pilot is too fatigued to fly safely. Computer programs such as the FlyAwake© and Fatigue Avoidance Scheduling Tool (FAST™) have been widely used in various military settings, incorporating information about recent sleep and work history to estimate a pilot's readiness-to-fly. Although these tools can be quickly and easily

used, they fail to take into consideration individual differences that may further influence a pilot's susceptibility to the effects of sleep loss. Further, research from Caldwell and colleagues (2005) indicated that those with lower levels of global activation as measured by a functional magnetic resonance imaging (fMRI) device tend to be more susceptible to the negative effects of sleep loss, but fMRI scans are far too expensive, complicated, and time-consuming to make them a practical readiness-to-fly assessment tool. Thus, the validation of a simple yet effective tool to accurately evaluate and predict a pilot's level of fatigue or general readiness to fly is still necessary.

Previous research has attempted to validate the use of physiologic and cognitive measures in detecting impairment due to fatigue (Chandler, Arnold, Phillips, Lojewski, & Horning, 2010). Using a 25-hour continuous wakefulness paradigm, the study evaluated participants on occulometric (PMI FIT 2000) and cognitive (Flight Fit) measures in conjunction with performance on the Psychomotor Vigilance Task (PVT), flight simulation, and working memory. Results of this study revealed that fatigue could be effectively gauged using both the PMI FIT and Flight Fit measures. Further, the authors concluded that these tools could also be used to distinguish between individual differences in susceptibility to fatigue more accurately than the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE™) model currently in use within the military services.

Although the findings from the Chandler et al. (2010) experiment do support the usage of the PMI FIT 2000 and Flight Fit in assessing an individual's susceptibility to fatigue and readiness-to-fly, that study was conducted with an acute sleep deprivation scenario, whereas fatigue associated with sustained operations is typically due to chronic sleep restriction (Caldwell, Chandler, & Hartzler, 2012). Before the PMI FIT and Flight Fit tools can be recommended for military implementation, it is necessary to ensure that they are accurate in evaluating fatigue and individual differences in fatigue susceptibility for chronic as well as acute sleep loss. Since individuals can differ widely in their susceptibility to fatigue due to sleep loss (Van Dongen, Baynard, Maislin, & Dinges, 2004; Van Dongen, Caldwell, & Caldwell, 2006), the validation of a tool which incorporates these individual differences would prevent those most susceptible to fatigue from being assigned duties at times when they should not be expected to perform at their best. Conversely, the same information could be used to identify personnel who are most fatigue resistant and thereby predict who would perform best during late night or sustained operations.

## METHOD

### Participants

Twenty-nine participants from the Naval Aviation Preflight Indoctrination program were recruited for participation in the present experiment. Five of these participants dropped out before completion of the study. Descriptive statistics for the participants who completed the study are presented in Table 1. To ensure participant safety, the study protocol was reviewed and approved by the Naval Aerospace Medical Research Laboratory (NAMRL) Institutional Review Board (IRB), in compliance with all applicable Federal regulations governing the protection of human subjects.

No specific groups were excluded from participation in this study. However, certain factors identified via a medical history form served to exclude individual participants due to their potential confounding effects (Killgore, Grugle, Reichart, Killgore, & Balkin, 2009). These include excessive alcohol use within the previous 48 hours (>3 drinks), greater than 400mg of routine daily caffeine consumption, habitual use of tobacco products within the previous six months, and history of significant medical, neurological, psychiatric, or sleep-related problems. The scientific literature does not provide clear evidence of the risk to pregnant women from limited sleep restriction; however, as a precautionary measure, and due to possible confounding effects, pregnant women were excluded from participation.

*Table 1*. Descriptive Statistics of Participants who Completed the Study

| | Age (years) | | Height (in) | | Weight (lbs) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Male (n = 21) | 19.9 | 2.7 | 69.9 | 2.8 | 163.6 | 24.3 |
| Female (n = 3) | 19.0 | 0.0 | 65.0 | 1.0 | 146.3 | 13.1 |
| Total | 19.8 | 2.5 | 69.3 | 3.1 | 161.5 | 23.7 |

**Fatigue Assessments**

***Revised NEO Personality Inventory***. The Revised NEO Personality Inventory (NEO-PI-R) is a widely used instrument for the assessment of personality functioning (Costa & McCrae, 1992). The inventory consists of 240 items answered on a five-point scale, ranging from "strongly disagree" to "strongly agree". The five domains measured are Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each domain is further subdivided into six facets that measure specific features of the primary personality factor. In their study, Killgore and colleagues (2007) determined that participants with higher Extraversion traits were more susceptible to fatigue in 77 hrs of wakefulness. In particular, the facets Gregariousness (E2) and Activity (E4) were the most sensitive to individual differences in fatigue susceptibility and thus will be the primary measures used for the present study. This inventory required approximately 40 minutes to complete and was only administered once during the Baseline Phase.

***Psychomotor Vigilance Task***. The Psychomotor Vigilance Task (PVT-192, Ambulatory Monitoring Inc., Ardsley, NY) is the gold standard instrument for assessment of reaction time and attention during periods of sleep loss (Balkin et al., 2004; Dinges et al., 1997). The task is completed using a small, battery-powered, hand-held device with two buttons and a small screen displaying the stimulus, numbers counted up in milliseconds. Participants are instructed to press the right button as soon as they notice numbers displayed on the screen. The numbers on the screen continue to count up either until the participant responds or until 1 minute (60,000ms) has passed. Lapses in performance are defined as any response time greater than 500ms.

***PMI FIT 2000.*** The PMI FIT 2000 (PMI) uses eye-tracking and pupillometry to identify impaired physiological states due to fatigue and other factors, such as alcohol or drug use. By comparing an individual's present state as measured on four pupillometric variables (Saccadic Velocity, Pupil Diameter, Pupil Constriction Amplitude, and Pupil Constriction Latency) with the same individual's baseline data, the system makes an evaluation of the individual's level of impairment. The system also compiles this data into a FIT Index, which is designed to be a

comprehensive estimate of the individual's current state. Each trial required approximately 30 seconds to complete.

***Flight Simulation (X-Plane 9)***. Simulated flight performance (FS) was measured using the X-Plane flight simulator (Laminar Research, Columbia, SC). Since fatigue impairs basic attentional processes, participants were given a simple flight profile, with instructions to fly "straight and level" due North at 140 knots and at an altitude of 2,000ft. Participants' performance was quantified by their ability to adhere to these parameters. The test required 20 minutes to complete.

***FlightFit***. The FlightFit (FF) neuropsychological test battery is an abbreviated version of the standard 30-minute CogniFit assessment battery. The tests measure cognitive performance on various components of mental work load sensitive to the effects of fatigue. The measures included a test of short term memory capacity, ability to focus attention, and visual scanning, as well as divided and shifting attention. This test required 10 minutes to complete.

***Dual n-back***. The *n*-back (NB) task is a measure of executive functioning consisting of a computer-based test in which the participant is presented with a sequence of stimuli (shapes, letters, numbers, or sounds) one at a time and is then required to recall the *n*-th stimulus back from the currently presented stimulus. For example, in a two-back task, if the participant was presented with the number string **1**, 3, **1**, 7, **5**, 4, **9**, 2, **9**, they would be required to indicate if the numbers highlighted in red (for the sake of illustration) match; that is, they were required to determine if the second number back matches the current number at every second progression. This protocol employed a form of the task called the dual *n*-back task, as described by Jaeggi, Buschkuehl, Jonides, and Perrig (2008). In the dual *n*-back, participants are presented with a series of shapes and sounds simultaneously. The specific *n* is adaptive and determined by the participant's performance as they progress through each trial. The task took approximately 20 minutes to complete.

***Stanford Sleepiness Scale***. The Stanford Sleepiness Scale (SSS) was included to assess participants' subjective sleepiness (Hoddes, Dement, & Zarcone, 1972). The SSS asks participants to indicate their level of sleepiness on a seven-point scale, from "1 - Feeling active, vital, alert, or wide awake" to "7 - No longer fighting sleep, sleep onset soon; having dream-like thoughts". There is also a means to denote if the participant is asleep, with the score of "X". The SSS is a widely used, easy-to-administer paper-and-pencil measure and has demonstrated excellent sensitivity to the effects of fatigue (Balkin et al., 2004). Participants were asked to indicate their level of sleepiness once during both of the training and baseline days and completion took less than a minute.

***Profile of Mood States - Brief***. The Profile of Mood States (POMS) is an assessment of transient, fluctuating active mood states (McNair, Lorr, & Droppleman, 1981). The survey was an abbreviated (30 items) version of the POMS standard (65 items). The survey measures Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, and Confusion-Bewilderment constructs. Items are measured on a five-point scale from "1 - Not at all' to "5 - Extremely." Additionally, the scores from these six factors are compounded to create the Total Mood Disturbance score. This is calculated by subtracting the Vigor-Activity score

from the sum of the scores of the other five facets. Each assessment requires approximately 5 minutes to complete.

***Fatigue Avoidance Scheduling Tool***. The Fatigue Avoidance Scheduling Tool (FAST; Nova Scientific Corporation, Fairborn, OH) is a computer program which was designed to predict any changes in performance due to sleep loss. This program was primarily intended to improve scheduling practices for operational aviation crews by taking into consideration the individual's recent work and rest schedules, and then predicting the extent to which fatigue might impact future work performance. Designed to reduce fatigue and fatigue-related errors and mishaps, predictions of effectiveness made by FAST are based on the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE™) model as well as results of numerous operational and laboratory findings (e.g., Hursh et al., 2004). For the present study, predictions of performance were calculated based on participants' individual sleep/wake patterns as recorded by the actigraph watch.

***University of Pennsylvania Smell Identification Test***. The University of Pennsylvania Smell Identification Test (UPSIT) consists of four self-administered booklets, each containing 10 different 'scratch & sniff' microencapsulated odor strips (Doty et al., 1995). A study by Killgore and colleagues (2010), determined that participants whose odor identification ability declined with continued wakefulness also demonstrated a decline in performance on executive functioning tasks while fatigued. Participants were assigned all four of the 10 item booklets, for a total of 40 different odors to identify. The test required approximately 15 minutes to complete, and was only administered at baseline and during the final test battery the following Thursday.

***Voice Analysis***. Voice analysis (VA) has shown promise in real-time fatigue monitoring (Greeley et al., 2007), especially for phrases involving hard "p" and "t" sounds. Participants were asked to read aloud five phrases commonly used by pilots. Exact phrases were: "I have the controls"; "Fuel: where is it supposed to be?"; "Requesting vectors"; "Traffic: high, low; factor, no factor"; and "Check gear, down and locked". Voice data was recorded and analyzed for systematic changes concurrent with the accumulation of sleep debt. The task took approximately 5 minutes to complete. Owing to the nature of this data, the statistical analyses required to examine the data were very different from those used for the other measures included in this study and thus will be described in a separate report.

**Design**

This experiment utilized a repeated measures design intended to further validate occulometric and cognitive individualized fatigue-detection technologies as potential readiness-to-fly tools, and to assess a wide array of potential individual fatigue detection approaches. Tests such as psychomotor vigilance, standardized flight simulator performance, personality and mood states, executive functioning, and odor identification were administered. To ensure sufficient levels of sleep debt, participants were permitted 4 hours sleep within each 24hour period over the course of 4 days. The experiment consisted of two phases, (1) the training/baseline phase and (2) the experimental, sleep restriction phase.

***Training/Baseline Phase***. Up to four volunteers were recruited from the Naval API student pool on the Thursday morning prior to each week of the study. After informed consent

was obtained on Thursday, participants completed the training session which included practice on all of the above measures except for the NEO-PI-R, UPSIT and FAST. Although data recorded this day was not included in any of the analyses, these test administrations were used to ensure that participants understood the instructions and were comfortable with the testing environment. During this same visit to the lab, participants were also outfitted with actigraph watches, which were used to monitor participants' sleep and wake patterns throughout the duration of the study.

When participants returned to the lab on Friday, they completed the same test session again and this data was used as the baseline measure. Participants also completed the NEO-PI-R and UPSIT on this day. Each day required approximately 2.5 hours of participation, for a total of 5 hours for this phase of the study.

***Experimental/Sleep restriction phase***. Upon completion of the Friday afternoon baseline data collection, participants were released with instructions to sleep according to their normal schedules and report to the Bachelor Officers' Quarters (BOQ) at 1600 Sunday evening. Participants were told to sleep from 0200 – 0600 hrs Monday morning and to then report to the Naval Aeromedical Research Laboratory (NAMRL) at 0700. To ensure participant safety throughout the sleep restriction phase of the study and prevent them from needing to drive, participants were required to sleep five nights (Sunday through Thursday) in the BOQ, located two blocks from NAMRL onboard Naval Air Station Pensacola. At the end of each experimental day, investigators debriefed the participants and escorted them back to the BOQ at approximately 0100.

Compliance to sleep time and amount was gauged by actigraphy data and noncompliance (i.e., sleeping 30 minutes or more over the 4 hours they were instructed to sleep) resulted in the participant's elimination from the experiment. Evaluation of eligibility for continuation was conducted via actigraph inspection from 0700 – 0730 each morning of the experimental phase. Also, participants were reminded of the daily schedule for the sleep restriction phase of the study (having initially been briefed during the informed consent session Thursday morning). Beginning at 0730 Monday participants were assessed on PMI, PVT, FS, VA, FF, NB, SSS, and POMS.

Task 1. 1 trial of PMI (1 min)
Task 2. 1 trial of PVT (10 min)
Task 3. 1 trial of FS (15 min)
Task 4. 1 trial of VA (5 min)
Task 5. 1 trial of FF (10 min)
Task 6. 1 trial of NB (20 min)
Task 7. 1 trial of SSS (1 min)
Task 8. 1 trial of POMS (5 min)

Testing was conducted at 0730, 1130, 1530, 1930, and 2330 each day of the experiment and each session lasted approximately 65-70 minutes. The testing blocks conducted at 0730, 1530, and 2330 included all eight of the tasks listed above, whereas the POMS (#8) was excluded from the testing blocks conducted at 1130 and 1930. Participants remained on the

premises between testing blocks, and were provided snacks, meals and beverages which did not interfere with or otherwise affect the experiment. Upon completion of the final trial on Thursday, participants were debriefed and driven to the BOQ, with instructions to obtain adequate sleep prior to check out. Late check-out provisions were prearranged with the BOQ. A table displaying the testing and sleep schedule is shown below.

*Table 2*. Schedule for Sleep and Data Collection during the Experimental Phase

| | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** |
|---|---|---|---|---|---|
| 0100 | | | | | Recovery Sleep |
| 0200 | Sleep Restriction Night 1 | Sleep Restriction Night 2 | Sleep Restriction Night 3 | Sleep Restriction Night 4 | |
| 0300 | | | | | |
| 0400 | | | | | |
| 0500 | | | | | |
| 0600 | | | | | |
| 0700 | 0730 - Test Session 1 | 0730 - Test Session 6 | 0730 - Test Session 11 | 0730 - Test Session 16 | |
| 0800 | | | | | |
| 0900 | | | | | |
| 1000 | | | | | |
| 1100 | 1130 - Test Session 2 | 1130 - Test Session 7 | 1130 - Test Session 12 | 1130 - Test Session 17 | |
| 1200 | | | | | |
| 1300 | | | | | |
| 1400 | | | | | |
| 1500 | 1530 - Test Session 3 | 1530 - Test Session 8 | 1530 - Test Session 13 | 1530 - Test Session 18 | |
| 1600 | | | | | |
| 1700 | | | | | |
| 1800 | | | | | |
| 1900 | 1930 - Test Session 4 | 1930 - Test Session 9 | 1930 - Test Session 14 | 1930 - Test Session 19 | |
| 2000 | | | | | |
| 2100 | | | | | |
| 2200 | | | | | |
| 2300 | 2330 - Test Session 5 | 2330 - Test Session 10 | 2330 - Test Session 15 | 2330 - Test Session 20 | |
| 2400 | | | | | |

## ANALYSES & RESULTS

### Overview

The data were analyzed in three stages which were intended to examine both group and individual differences on a number of tasks in response to chronic sleep restriction. To determine whether any of the measures captured significant change in performance during the sleep restriction phase, Stage 1 included a series of Repeated Measures Analyses of Variance (ANOVAs) to identify those variables which were significantly affected by chronic sleep restriction. Any significant variables from the Stage 1 analyses were then included in Stage 2, a series of Hierarchical Linear Model (HLM) analyses which were conducted to tease apart

individual differences not evident in the group level analyses, as well as to predict performance deterioration due to fatigue. For Stage 3, the significant predictor variables from Stage 2 were utilized in a series of enter-method linear regression analyses to develop predictive algorithms to estimate performance deterioration when fatigued. The analyses for Stages 1 and 3 were performed using SPSS version 16.0 for Windows (SPSS Inc., Chicago, IL). The analyses for Stage 2 were conducted using HLM version 6.02 for Windows (Scientific Software International, Inc., Skokie, IL). For all three stages of analyses reported herein, a criterion for statistical significance was set to a value of $p \leq 0.05$.

**Stage 1**

Data collected during the test session on Friday were used to establish baseline and were included in the following analyses. A series of repeated measures were conducted for each of the dependent variables collected over the baseline and subsequent 20 testing sessions (12 for the POMS) during the sleep restriction phase of the study.

*Psychomotor Vigilance Task (PVT).* Two variables from the PVT were analyzed: number of PVT lapses and mean SRRT, the mean reciprocal reaction time of the slowest 10% of responses. Results revealed significant fatigue effects for lapses and mean SRRT, as depicted by Table 3 and Figures 1 and 2. Specifically, over the course of the sleep restriction period of the study, the mean number of lapses increased and mean SRRT deteriorated. Post-hoc analyses indicated that number of lapses at baseline was significantly different from that of Trials 3 – 20. However, results for the mean SRRT indicated that performance on all 20 trials was significantly different from that recorded during baseline testing. Since the mean number of PVT lapses is known in the existing literature as a gold standard for measuring fatigue, this measure was included as the primary criterion variable in Stage 2 analyses.

*Table 3.* ANOVA results for PVT

| | F | df | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| PVT Lapses† | 9.499 | (4.934, 113.472) | .000* | 0.292 |
| Mean SRRT† | 18.801 | (7.913, 182.009) | .000* | 0.450 |

*Significant at the .05 level

†Greenhouse-Geisser correction used due to violation of sphericity

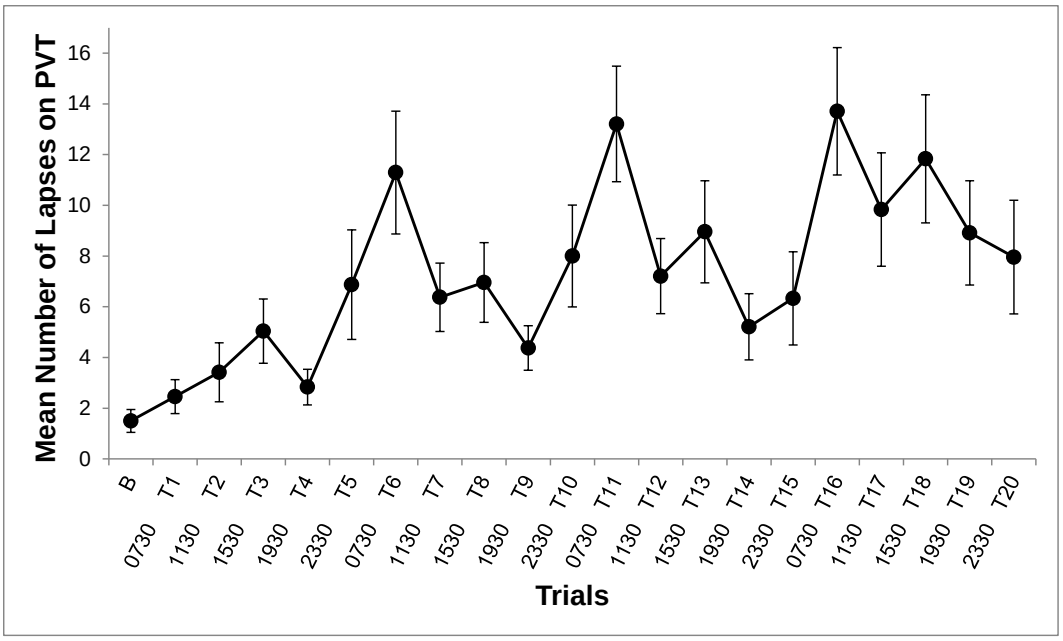


Figure 1. Mean PVT lapses at each test trial across time. Post-hoc analyses revealed significant differences between the baseline measure (B) and Trials 3 – 20.

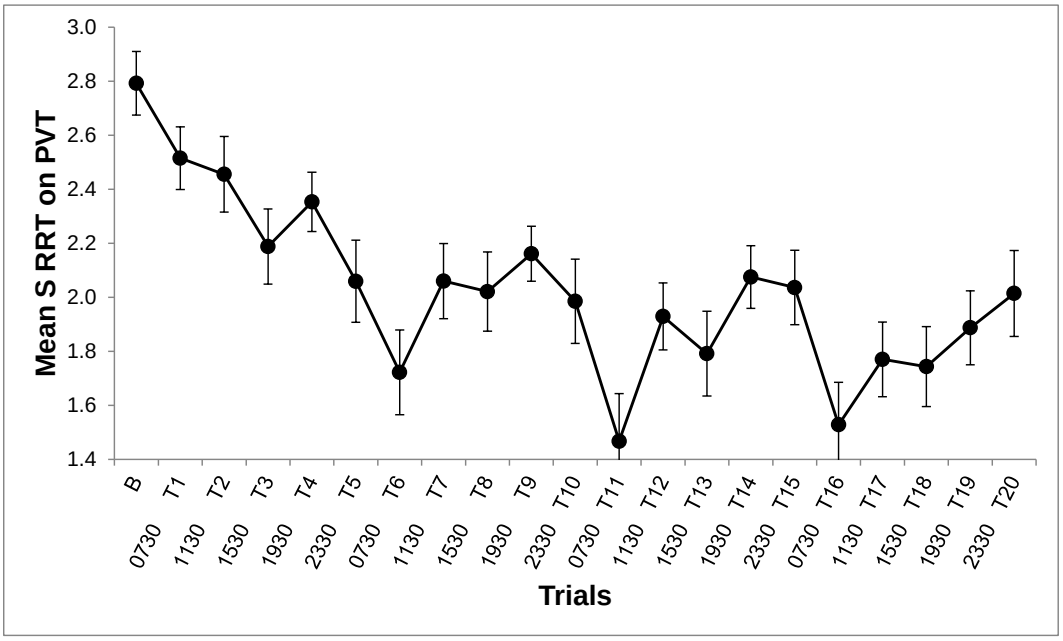


Figure 2. Mean reciprocal reaction time of the slowest 10% of responses (mean SRRT) for the PVT at each test trial across time. Post-hoc analyses revealed significant differences between the baseline measure (B) and Trials 1 – 20.

Additionally, two further analyses were conducted using PVT data to determine whether the data obtained in this study replicated the findings of Killgore and colleagues (2007) regarding increased fatigue susceptibility among participants who rate more highly on two of the NEO-PI-R Extraversion facets (Table 4). In the Killgore study, results indicated that after two consecutive nights of total sleep deprivation the Extraversion facets Gregariousness (E2) and

Activity (E4) were the only measures which were significantly correlated with participants' change in speed from baseline measures on the PVT. For these analyses, the percent change in speed from baseline was calculated as the average reciprocal reaction time (i.e., 1 / RT) times 100, then divided by the reciprocal of the baseline measure. The same transformation utilized by Killgore et al. was used in the present study to compute change in speed scores, which employed a chronic sleep restriction design rather than total sleep deprivation. For these analyses, neither the Gregariousness nor Activity facets predicted changes in PVT speed, though evident trends were similar to those reported by Killgore. As demonstrated in Figure 3, participants who were ranked as being the most gregarious (i.e., in the upper 25% on the facet) appeared to have the greatest deterioration in performance across the period of sleep restriction. Likewise, participants who were ranked in the highest 25% on the Activity facet also exhibited a growing impairment during the period of sleep loss to an extent greater than that of participants in the lower 25% and middle 50% of the facet (Figure 4).

*Table 4.* ANOVA results for PVT

| | F | df | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| PVT Speed + NEO-PI-R Gregariousness Rating† | 1.368 | (12.586, 132.153) | .185 | 0.115 |
| PVT Speed + NEO-PI-R Activity Rating† | 1.052 | (12.603, 132.330) | .406 | 0.091 |

*Significant at the .05 level

†Greenhouse-Geisser correction used due to violation of sphericity

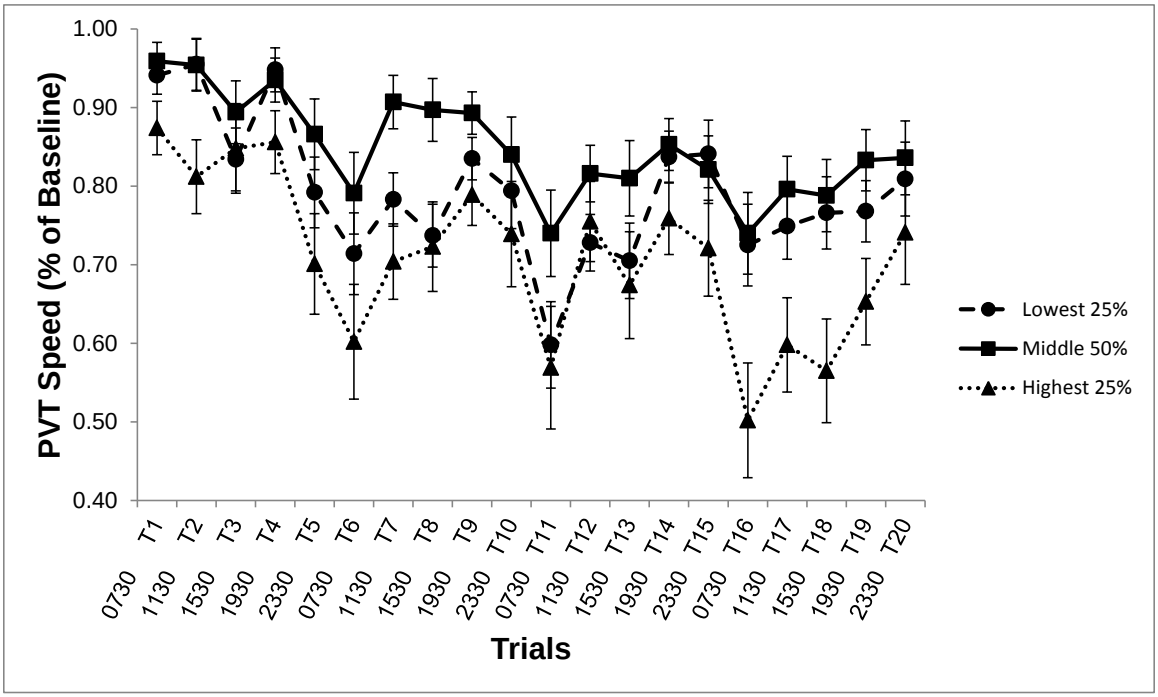


Figure 3. Changes in PVT speed at each test trial across time with participants grouped by their NEO-PI-R Gregariousness rating. No significant between-subjects effects were evident.

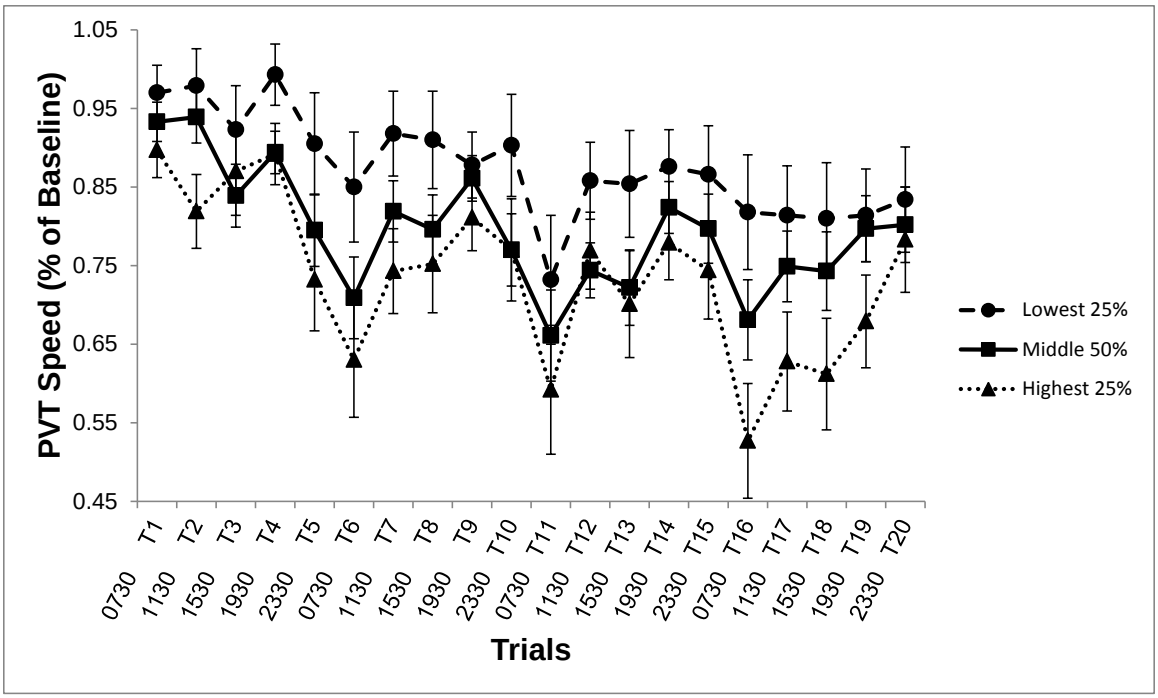


Figure 4. Changes in PVT speed at each test trial across time with participants grouped by their NEO-PI-R Activity rating. No significant between-subjects effects were evident.

*PMI Fit 2000.* The PMI FIT collects data on four oculometric measures: Pupil Constriction Latency, Pupil Constriction Amplitude, Pupil Diameter, and Saccadic Velocity. These four components are also combined to produce the overall FIT Index. Results from these measures are displayed in Table 5 and Figures 5 – 9 below. As shown, all subcomponents of the FIT Index exhibited significant changes in response to fatigue, though changes in the actual FIT Index were not significant. For the FIT Index, the average of the baseline measures was excluded from analysis because the mean value was more than 20 times greater than any of the mean scores recorded during the Experimental Phase (Figure 9).

*Table 5.* ANOVA results for PMI

| | F | df | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| Constriction Latency† | 2.636 | (9.926, 228.300) | .005* | 0.103 |
| Constriction Amplitude | 2.666 | (20, 460) | .000* | 0.104 |
| Pupil Diameter† | 5.324 | (8.642, 198.768) | .000* | 0.188 |
| Saccadic Velocity† | 3.871 | (9.002, 207.052) | .000* | 0.144 |
| FIT Index† | 1.657 | (6.581, 138.211) | .129 | 0.073 |

*Significant at the .05 level

†Greenhouse-Geisser correction used due to violation of sphericity

Post-hoc analyses revealed that for the subcomponent Pupil Constriction Latency, the average of the baseline measures was significantly different only from Trial 5. Conversely, Pupil Constriction Amplitude demonstrated significant deviations from baseline for Trials 1 – 3, 6, 11, and 16 (Figure 6).

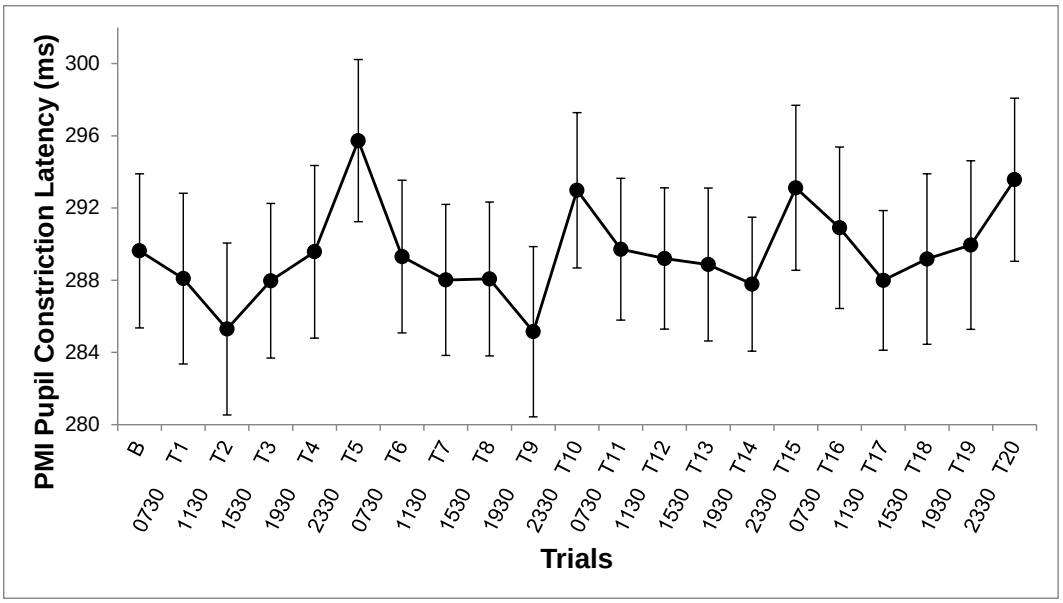


Figure 5. Mean PMI Pupil Constriction Latency in milliseconds at each test trial across time. Post-hoc analyses revealed significant differences between the averages of the baseline

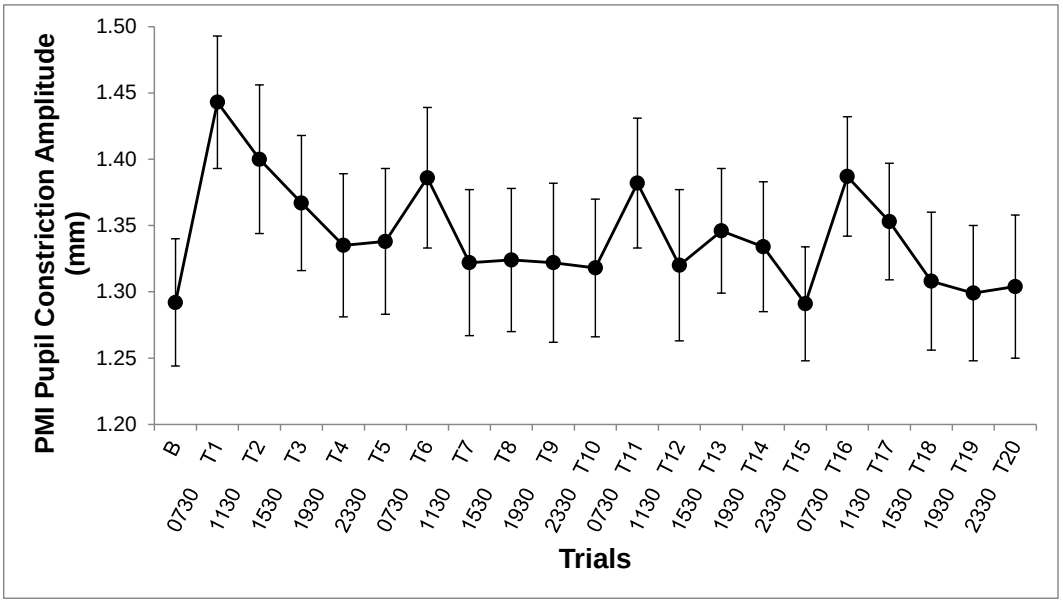


Figure 6. Mean PMI Pupil Constriction Amplitude in millimeters at each test trial across time. Post-hoc analyses revealed significant differences between the averages of the baseline measures (B) and Trials 1-3, 6, 11, and 16.

Additionally, two other subcomponents of the FIT demonstrated sensitivity to the effects of fatigue due to sleep restriction – Pupil Diameter and Saccadic Velocity. Specifically, Pupil Diameter measurements during the baseline testing were significantly different from measurements recorded during Trials 3, 4, and 14 – 20 (Figure 7). Additionally, post-hoc analyses indicated that Saccadic Velocity on the averaged baseline measures was significantly different from that seen on Trials 3, 5, 6, 8 – 12, 15 – 18 (Figure 8).

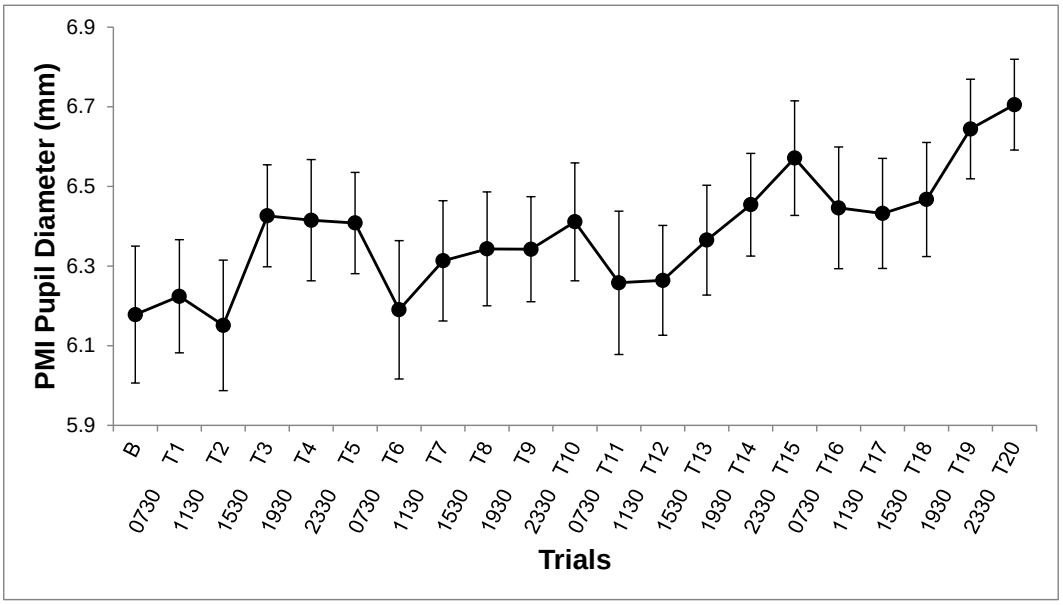


Figure 7. Mean PMI Pupil Diameter in millimeters at each test trial across time. Post-hoc analyses found significant differences between the average of the baseline measures (B) and Trials 3, 4, and 14 – 20.

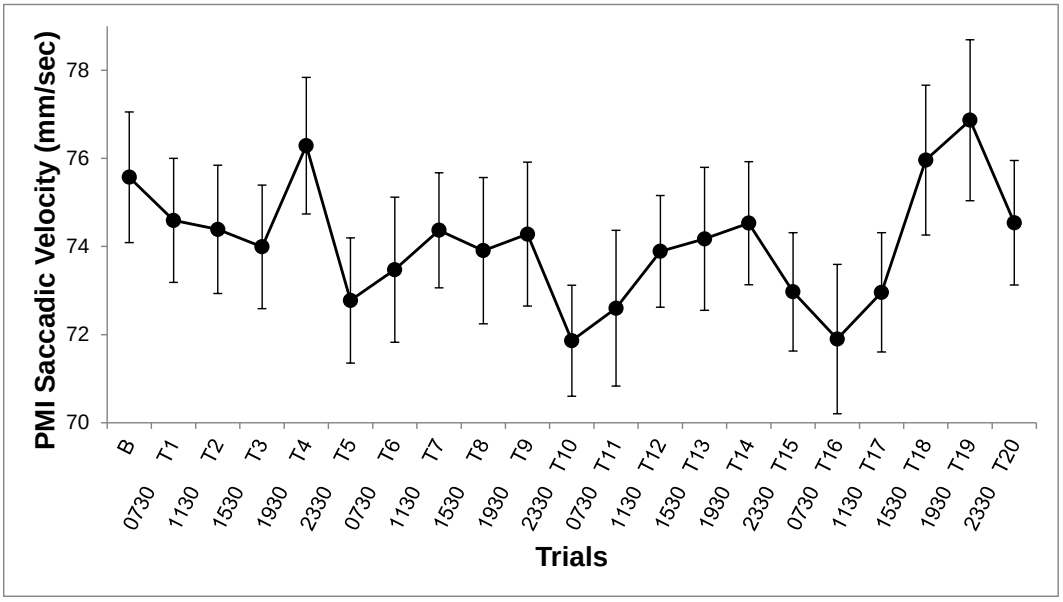


Figure 8. Mean PMI Saccadic Velocity in millimeters per second at each test trial across time. Post-hoc analyses revealed significant differences between the average of the baseline measures (B) and Trials 3, 5, 6, 8 – 12, 15 – 18.

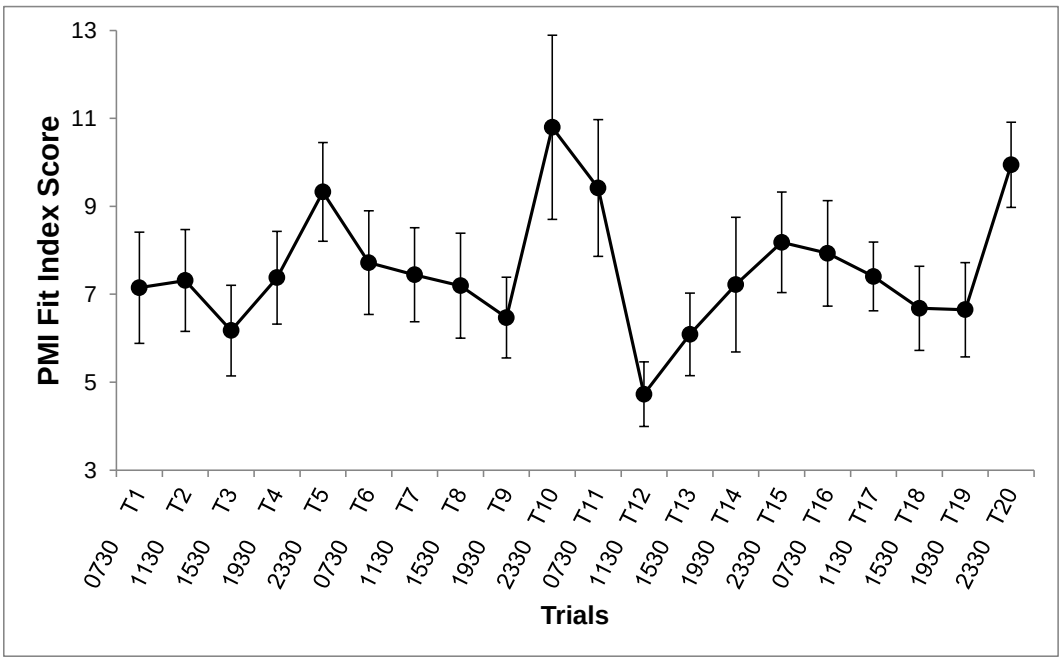


Figure 9. Mean PMI Fit Index Scores at each test trial across time which revealed no significant change in score across the Experimental Phase.

*Flight Simulator*. Deviations from the specified flight parameter goals for heading (due North), airspeed (140 knots), and altitude (2,000 ft) were calculated separately. Lapse times for each parameter were calculated as the number of seconds during a simulator trial that subjects deviated from the flight goal by greater than one standard deviation (determined at baseline). Total lapse time was the sum of lapse times for each flight parameter for each testing session. The analysis revealed dramatic and significant effects of the time of testing on total lapse time suggesting that total lapse time is sensitive to fatigue effects (see Table 6 and Figure 10). Post-hoc analyses indicated that the Total Lapse Time during the baseline testing session was significantly less than that of any of the 20 subsequent test sessions.

*Table 6*. ANOVA for Flight Simulator Total Lapse Time†

| F | df | *p* | $\eta_p^2$ |
|---|---|---|---|
| 6.263 | (7.903, 181.779) | .000* | 0.214 |

*Significant at the .05 level

†Greenhouse-Geisser correction used due to violation of sphericity

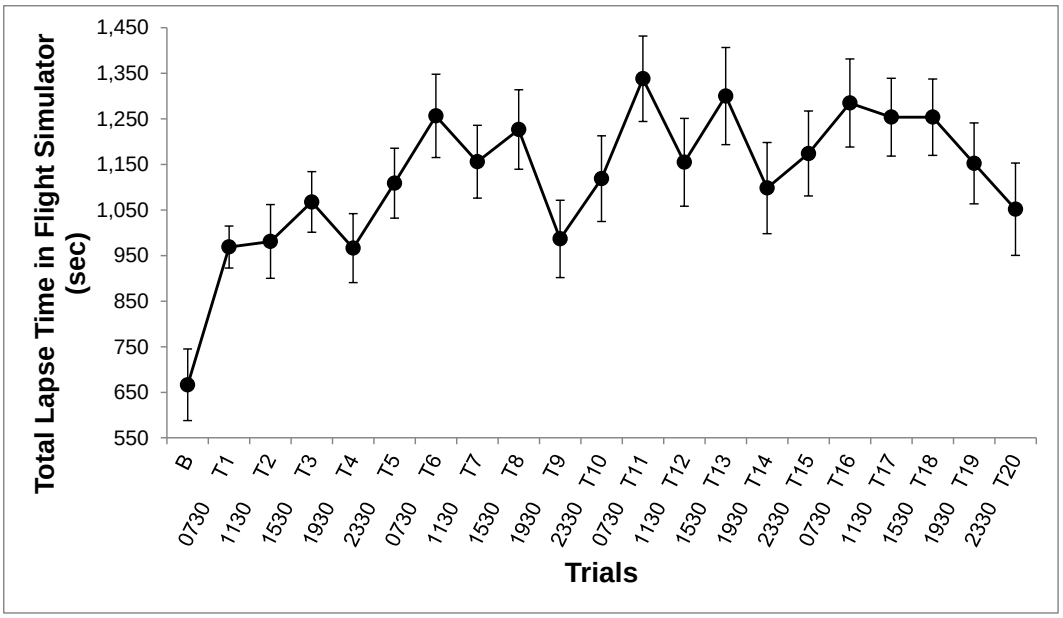


Figure 10. Flight Simulator Total Lapse Time in seconds across time. Post-hoc analyses revealed significant differences between the baseline test session (B) and Trials 1 – 20.

*Flight Fit*. Each of the 9 subscales for the Flight Fit test were analyzed to identify any changes in performance during the course of the Experimental Phase, and a summary of the results are displayed in Table 7 and Figures 11 – 19. For each of the subscales, data from four of the participants on Trial 12 was not recorded by the computer and thus this trial was removed from the analyses.

Repeated measures analyses indicated that performance on three of the measures changed significantly over the course of the Experimental Phase: Raw Response Time; Visual Scanning Response Time; and Divided Attention Response Time. Subsequent post-hoc analyses of these measures revealed significant differences between the baseline measure and trials recorded during the Experimental Phase. For Raw Response Time this difference was only significant for Trial 1, whereas for both Visual Scanning Response Time and Divided Attention Response time, the difference was significant for Trials 2 – 20. In general, the change observed among the response time measures indicates that, despite chronic sleep restriction, participants speed in completing the tasks increased. Though the three measures listed above were significant for Stage 1 analyses, the results suggest that the changes were due to practice effects rather than increasing fatigue and thus were not included in subsequent analyses.

*Table 7.* ANOVA results for Flight Fit

| | F | df | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| Raw Response Time | 2.678 | (19, 437) | .000* | 0.104 |
| Difference Score for Focus† | 0.989 | (10.10., 232.370) | .454 | 0.041 |
| Visual Scanning Accuracy† | 1.016 | (9.036, 207.831) | .429 | 0.042 |
| Visual Scanning Response Time† | 5.609 | (8.480, 195.034) | .000* | 0.196 |

| | | | | |
|---|---|---|---|---|
| Divided Attention Accuracy† | 1.330 | (9.381, 196.999) | .221 | 0.060 |
| Divided Attention Response Time† | 5.515 | (7.604, 159.676) | .000* | 0.208 |
| Shifting Attention Accuracy† | 1.226 | (3.741, 86.036) | .306 | 0.051 |
| Shifting Attention Response Time† | 1.828 | (9.303, 195.372) | .063 | 0.080 |
| Short Term Memory† | 1.393 | (8.022, 104.289) | .208 | 0.097 |

*Significant at the .05 level
†Geisser-Greenhouse correction used due to violation of sphericity

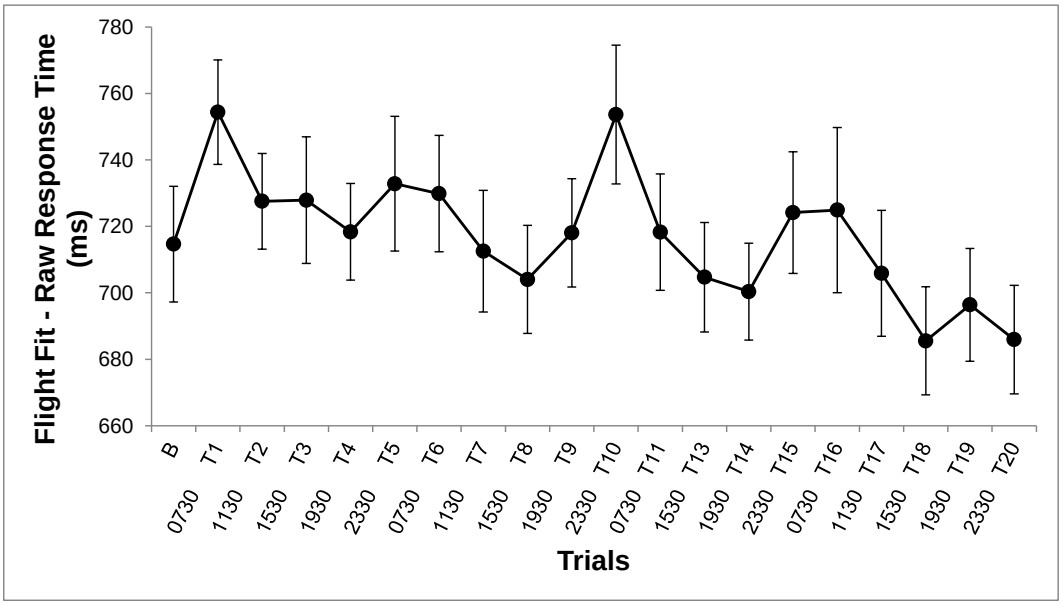


Figure 11. Flight Fit subscale, Raw Response Time in milliseconds across time. Post-hoc analyses revealed significant differences between the average of the baseline measures (B) and Trial 1.

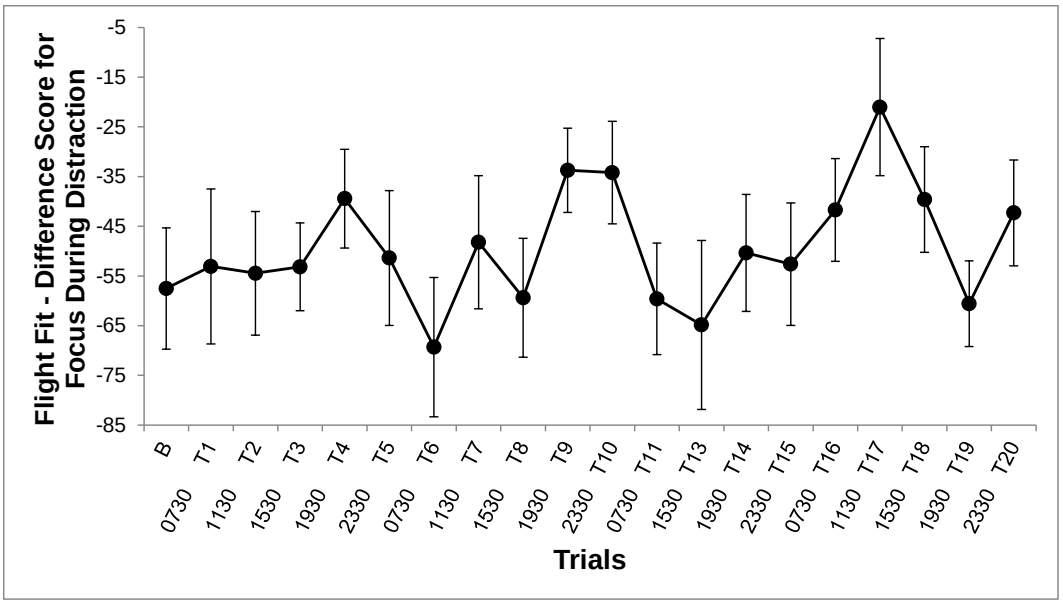


Figure 12. Flight Fit subscale, difference between response times for a two-part task which required focus in the presence of distractors. No significant effects of fatigue were detected.

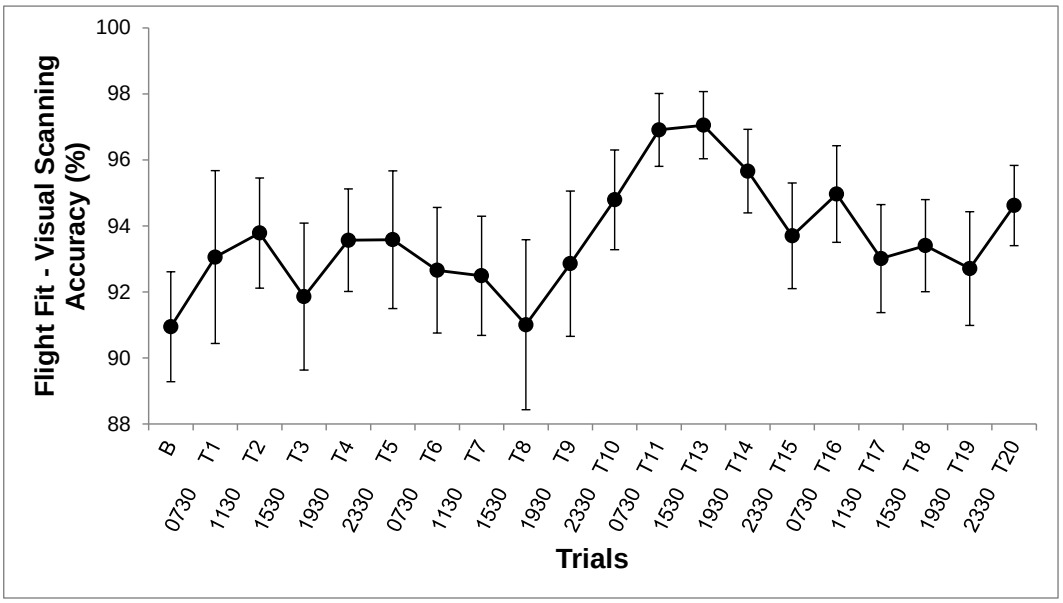


Figure 13. Flight Fit subscale, percent correct on Visual Scanning task. No significant effects of fatigue were detected.

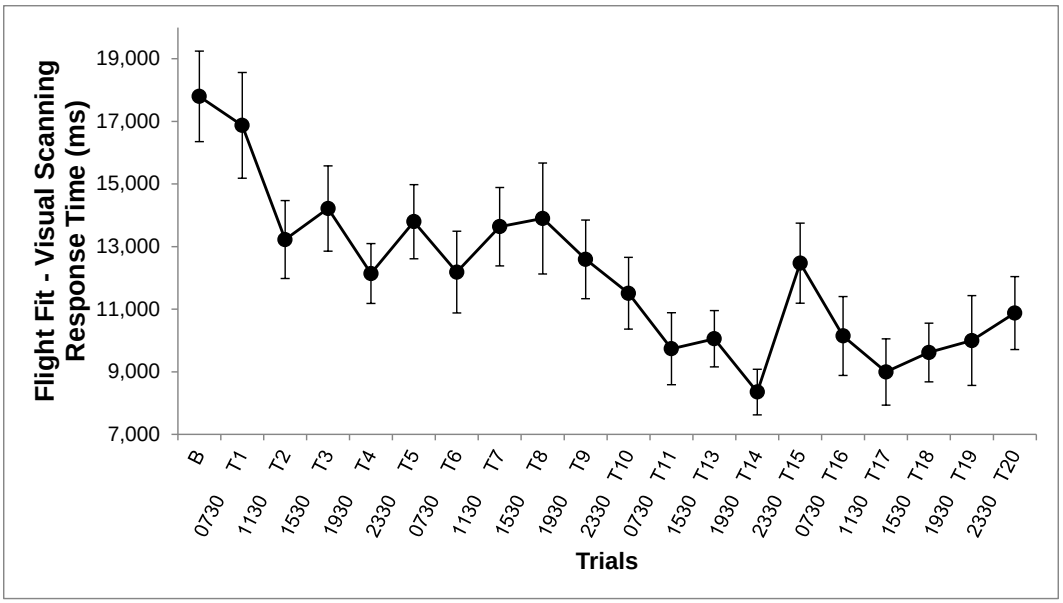


Figure 14. Flight Fit subscale, response time in milliseconds (msec) on Visual Scanning task. Post-hoc analyses revealed significant differences between the average baseline measures (B) and Trials 2 – 20.

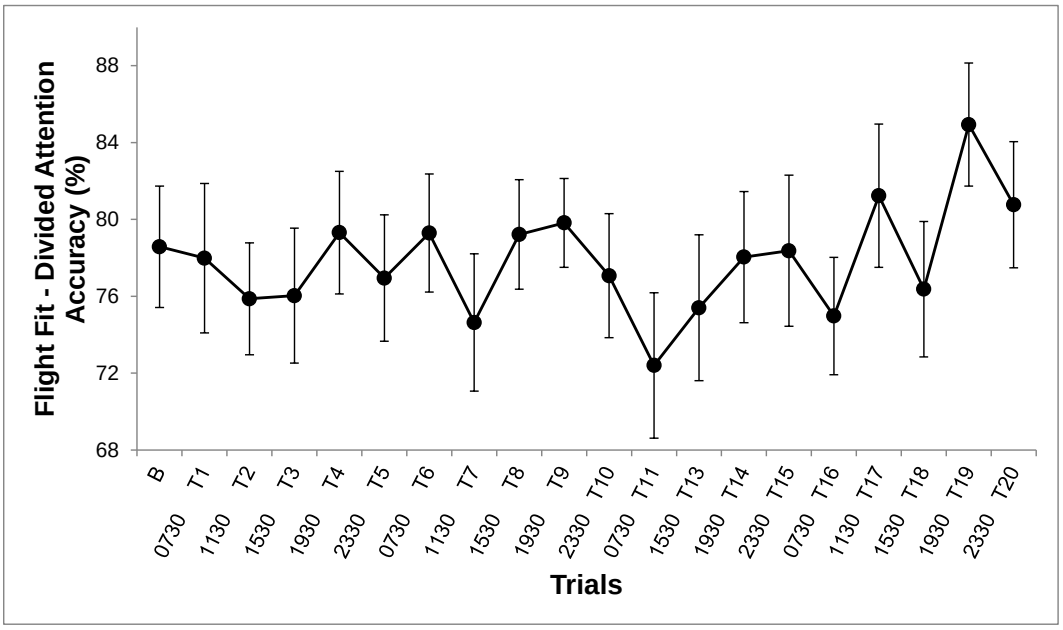


Figure 15. Flight Fit subscale, percent correct on Divided Attention task. No significant effects of fatigue were detected.

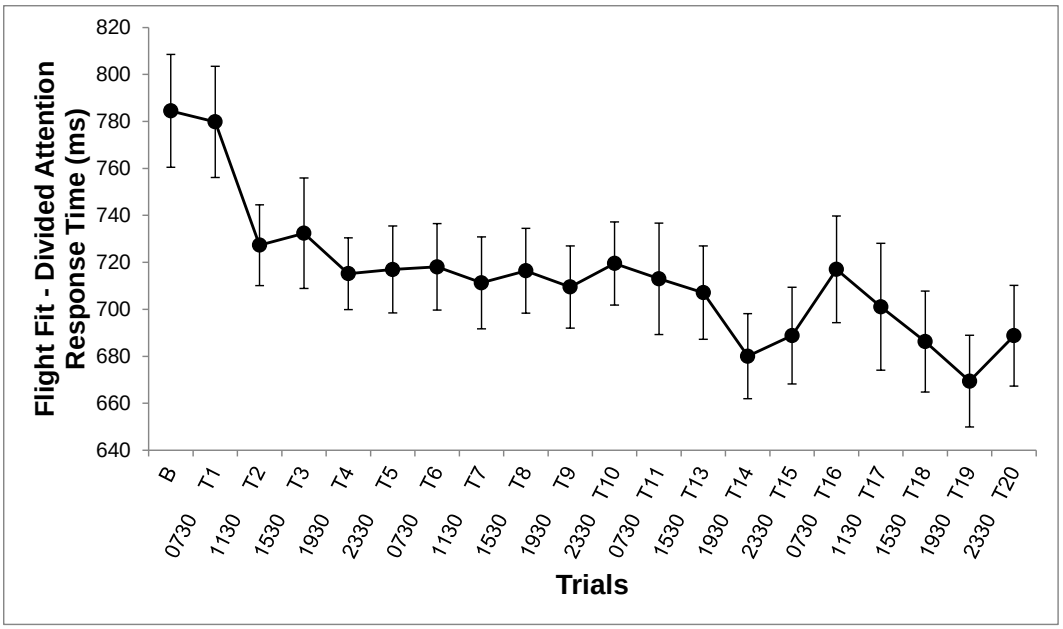


Figure 16. Flight Fit subscale, response time on Divided Attention task. Post-hoc analyses revealed significant differences between the average of the baseline measures (B) Trials 2 - 20.

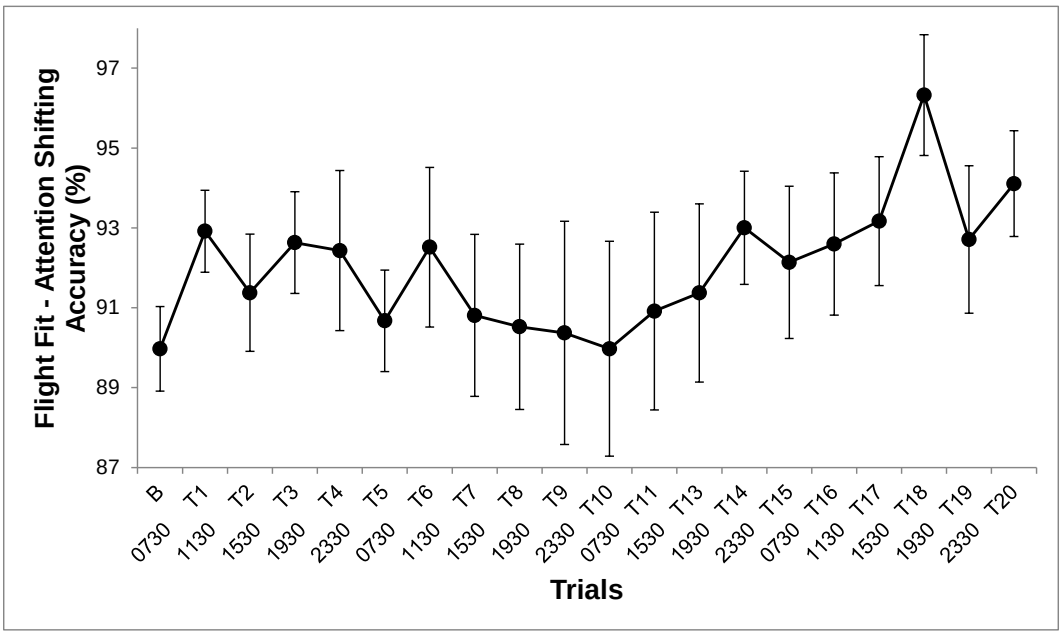


Figure 17. Flight Fit subscale, percent correct on Attention Shifting task. No significant effects of fatigue were detected.

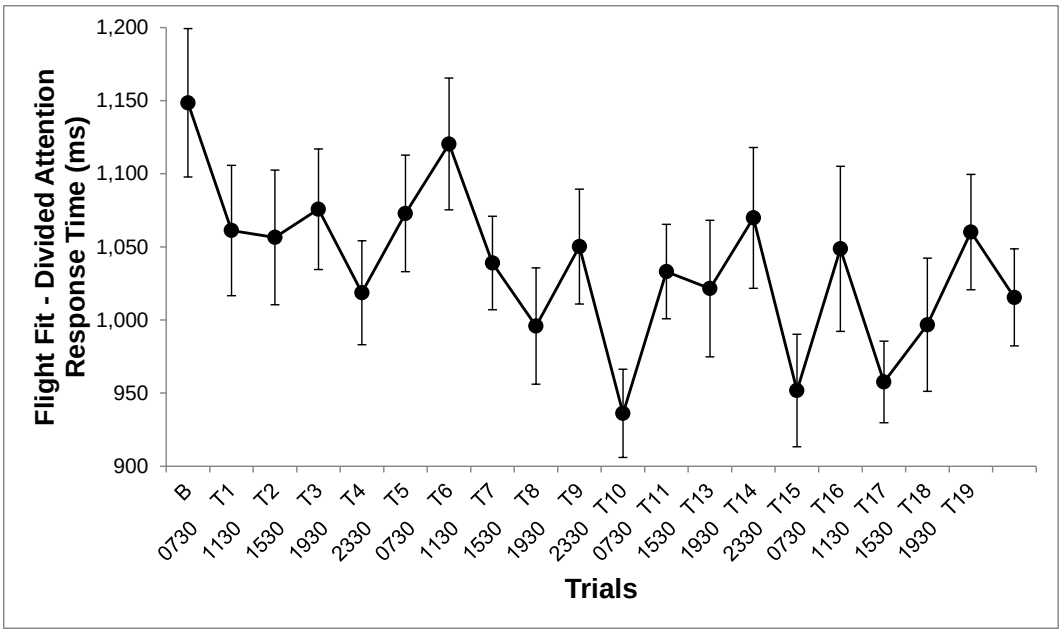


Figure 18. Flight Fit subscale, response time for Attention Shifting task. No significant effects of fatigue were detected.

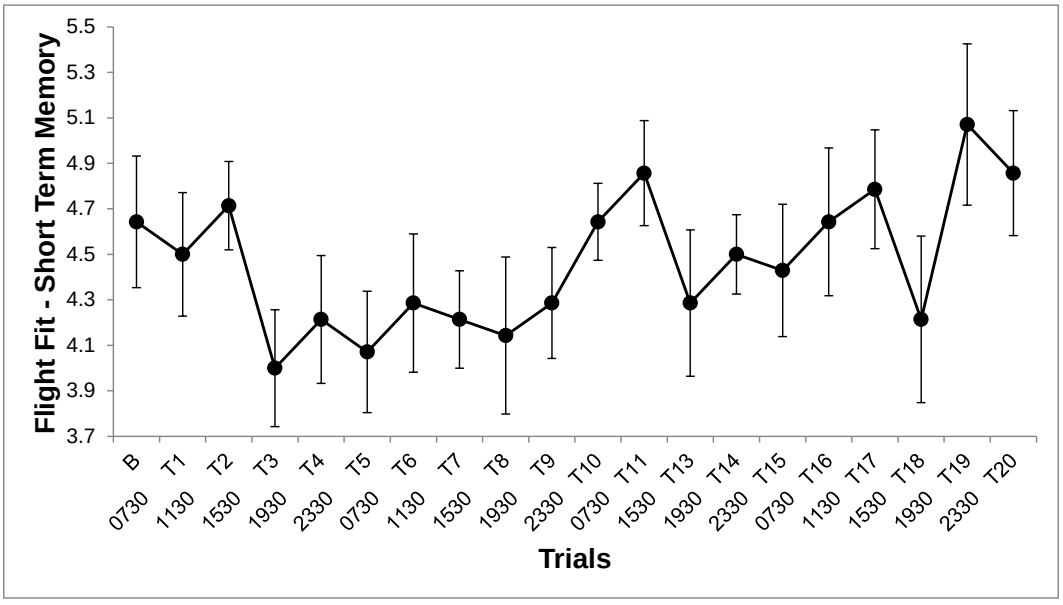


Figure 19. Flight Fit subscale, number memorized on the Short Term Memory (STM) task. No significant effects of fatigue were detected.

*Dual N-Back.* Results of the ANOVA for the dual *n*-back were significant (see Table 8), indicating a general improvement in performance throughout the course of the sleep restriction period. Post-hoc analyses revealed that baseline performance was significantly poorer from that of all trials during the Experimental Phase, suggesting that, despite sleep restriction, practice effects were evident throughout the study (Figure 20). Consequently, this measure was not included in any subsequent analyses.

*Table 8.* ANOVA for Dual *n*-Back†

| F | df | *p* | $\eta_p^2$ |
|---|---|---|---|
| 6.595 | (7.027, 161.626) | 0.000* | 0.223 |

*Significant at the .05 level
†Greenhouse-Geisser correction used due to violation of sphericity

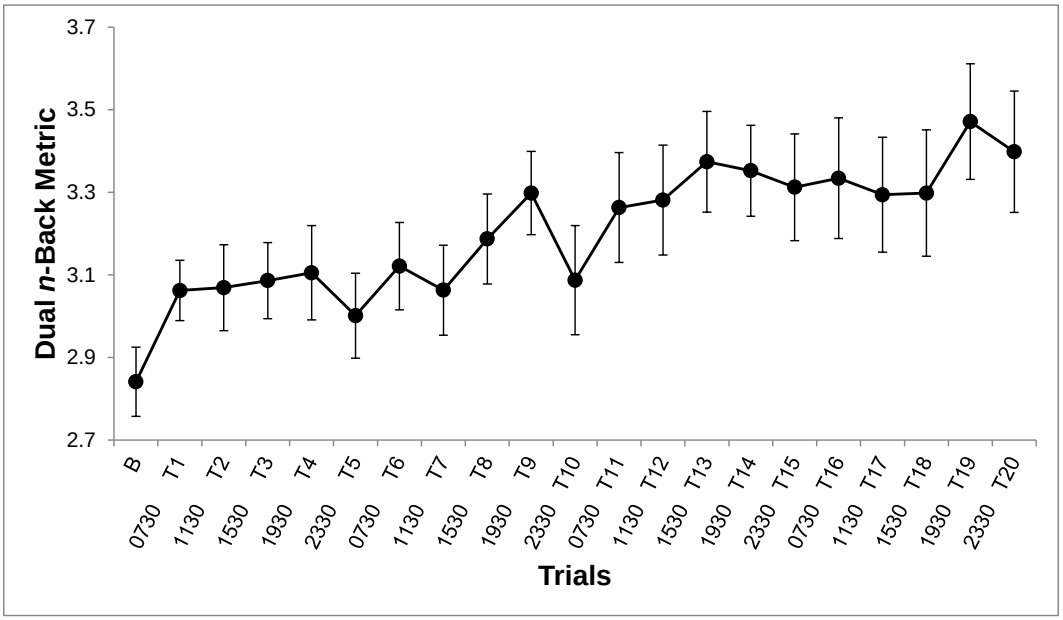


Figure 20. Dual N-back at each test trial across time. Post-hoc analyses revealed significant differences between the average of the baseline measures (B) and all trials during the Experimental Phase.

*Stanford Sleepiness Scale (SSS).* Results for the SSS scores indicate that there was a significant change in self-reported sleepiness with estimates gradually increasing across the Experimental Phase (see Table 9 and Figure 21). Post-hoc analyses indicated that subjective sleepiness during the baseline testing was significantly different from that reported on all 20 of the subsequent trials.

*Table 9.* ANOVA results for Stanford Sleepiness Scale†

| F | df | *p* | $\eta_p^2$ |
|---|---|---|---|
| 10.723 | (7.314, 160.919) | 0.000* | 0.328 |

*Significant at the .05 level
†Greenhouse-Geisser correction used due to violation of sphericity

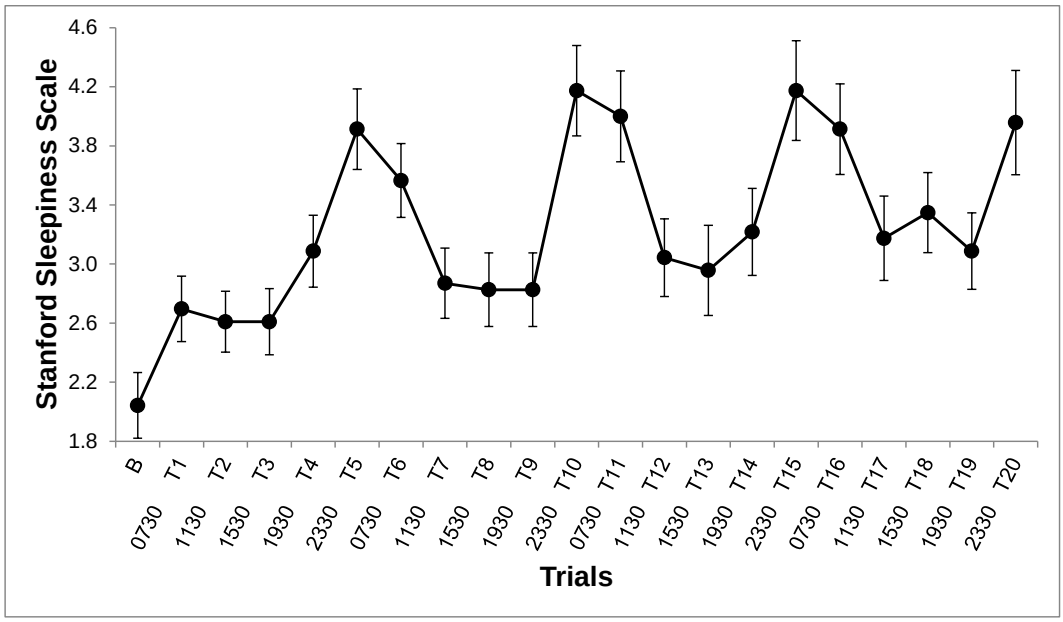


Figure 21. Stanford Sleepiness Scale at each test trial across time. Post-hoc analyses revealed that participant responses to the questionnaire during baseline testing was significantly different than responses all trials during the sleep restriction period.

*Profile of Mood States (POMS).* Participants were only asked to complete this questionnaire three times a day during the Experimental Phase, at the end of the 0730, 1530, and 2330 test sessions. The POMS questionnaire evaluates participants on six different mood components: Tension-Anxiety, Depression-Dejection, Anger-Hostility, Fatigue-Inertia, Confusion-Bewilderment, and Vigor-Activity. Additionally, participants' scores on these six components are used to calculate the Total Mood Disturbance rating by adding the values from the first five measures and then subtracting the score from the sixth. Results of these analyses are detailed below in Table 10 and Figures 22 – 28. Four of these measures, Tension-Anxiety, Depression-Dejection, Anger-Hostility, and Confusion-Bewilderment were not sensitive to any changes in fatigue levels as a result of sleep restriction. However, results from the remaining three components (Fatigue-Inertia, Vigor-Activity, and Total Mood Disturbance) indicated that these measures were sensitive to the changes in emotional state during the sleep restriction period.

*Table 10*. ANOVA results for Profile of Mood States

| | F | df | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| Tension-Anxiety† | 0.594 | (4.384, 100.823) | 0.683 | 0.025 |
| Depression-Dejection† | 0.946 | (4.711, 108.354) | 0.451 | 0.040 |
| Anger-Hostility† | 1.701 | (2.210, 50.822) | 0.190 | 0.069 |
| Confusion-Bewilderment† | 2.725 | (2.877, 66.178) | 0.053 | 0.106 |
| Vigor-Activity† | 17.772 | (3.655, 84.054) | 0.000* | 0.436 |
| Fatigue-Inertia† | 17.603 | (4.823, 110.920) | 0.000* | 0.434 |

| | | | | |
|---|---|---|---|---|
| Total Mood Disturbance† | 24.147 | (4.957, 114.011) | 0.000* | 0.512 |

*Significant at the .05 level
†Greenhouse-Geisser correction used due to violation of sphericity

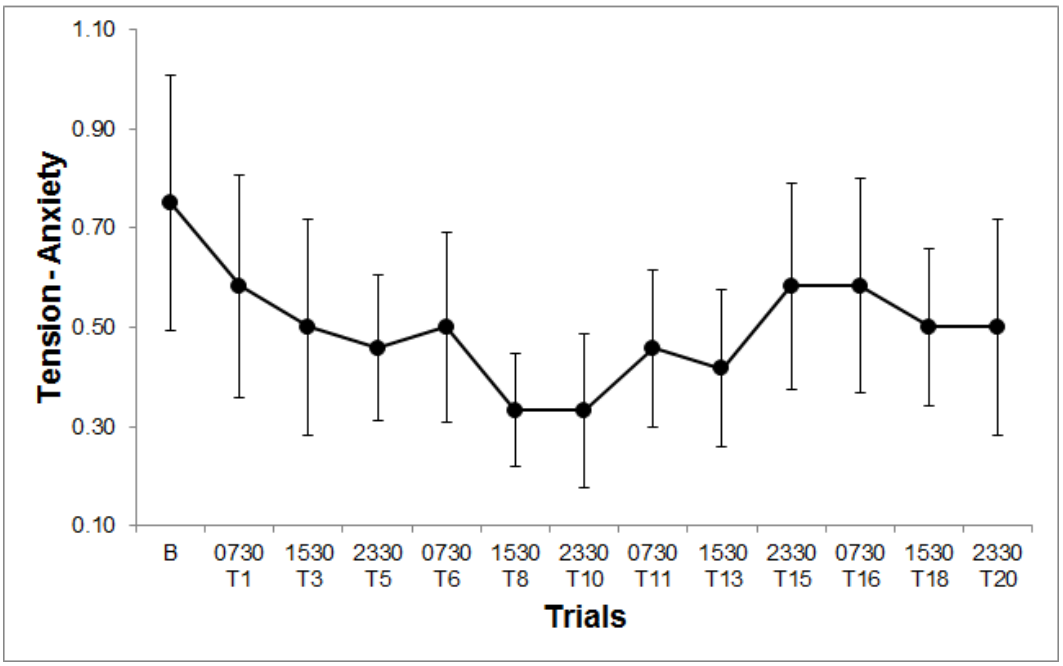


Figure 22. POMS Tension-Anxiety scores across the Experimental Phase. No significant fatigue effects were detected.

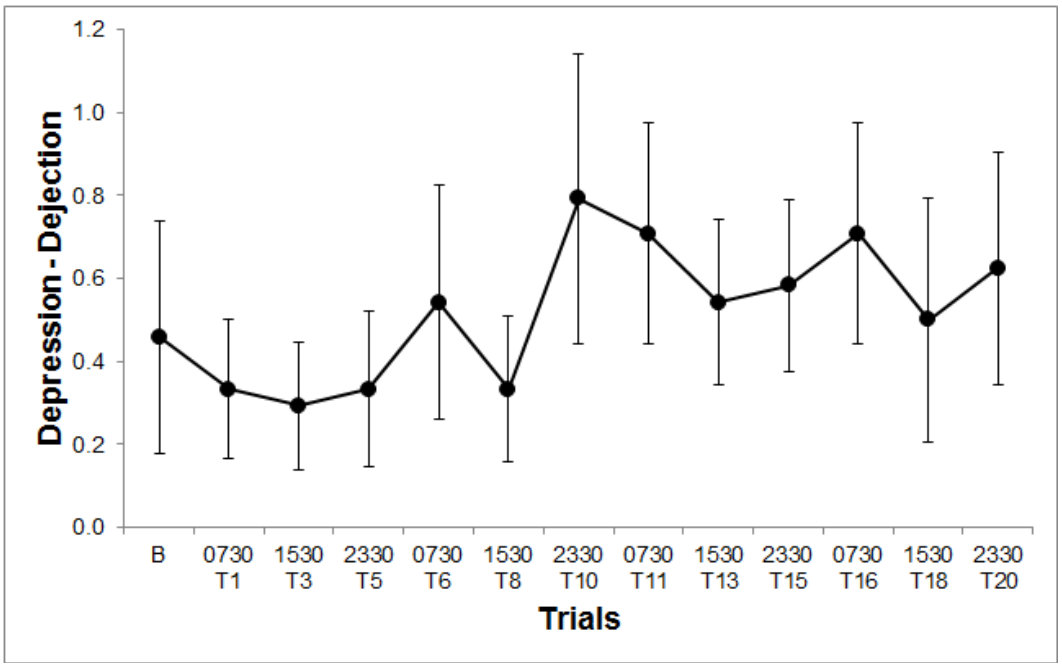


Figure 23. POMS Depression-Dejection scores across the Experimental Phase. No significant fatigue effects were detected.

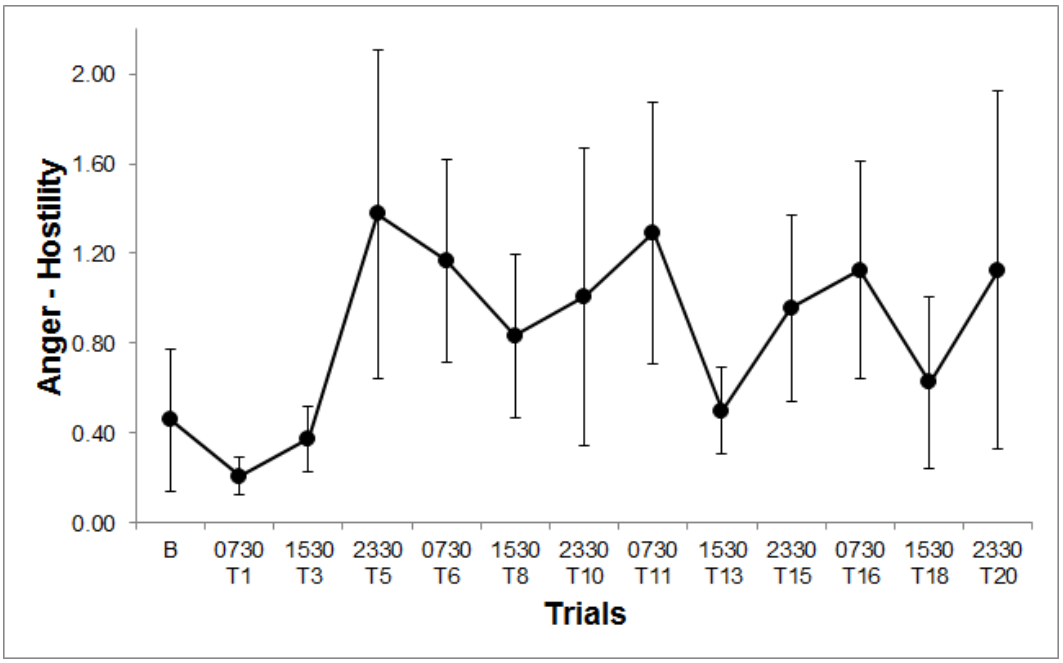


Figure 24. POMS Anger-Hostility scores across the Experimental Phase. No significant fatigue effects were detected.

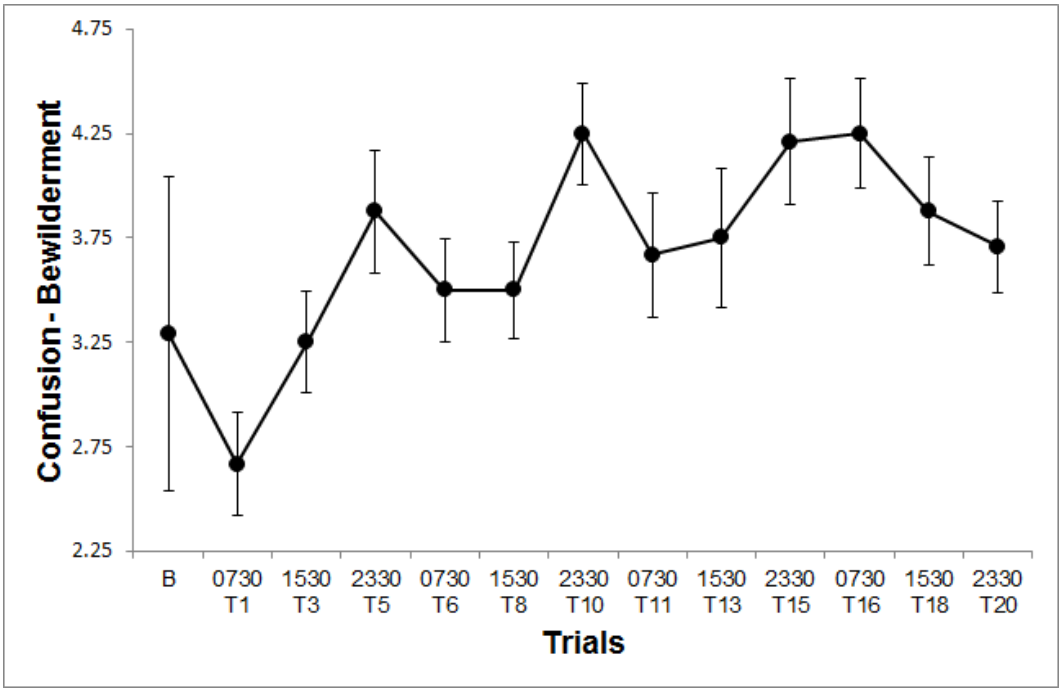


Figure 25. POMS Confusion-Bewilderment scores across the Experimental Phase. No significant fatigue effects were detected.

Vigor-Activity changed significantly over the course of the Experimental Phase and tended to decrease overall (Figure 26), whereas Fatigue-Inertia (Figure 27) and Total Mood Disturbance (Figure 28) also changed significantly, but the subjective ratings for these states generally increased. Thus, subjective assessments of vitality and cognitive clarity appear to be

the most sensitive to increasing levels of fatigue during sleep restriction and these measures were included in the Stage 2 analyses. Post-hoc analyses revealed significant differences between the baseline measure and some of the subsequent trials for all three of these components. Specifically, baseline scores on both Vigor-Activity and Total Mood Disturbance were significantly different from each of the 12 scores recorded during the sleep restriction phase, whereas baseline scores for Fatigue-Inertia differed significantly from Trials 3 – 20.

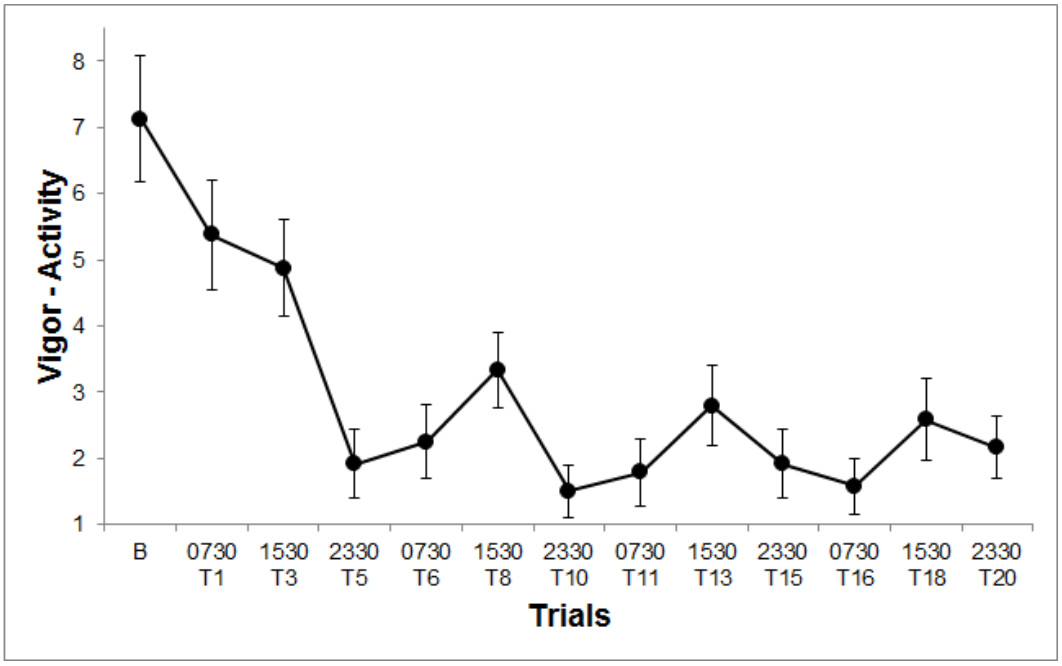


Figure 26. POMS Vigor-Activity scores across the Experimental Phase. Significant fatigue effects were evident, with responses during the baseline testing session being significantly different from responses during all subsequent trials.

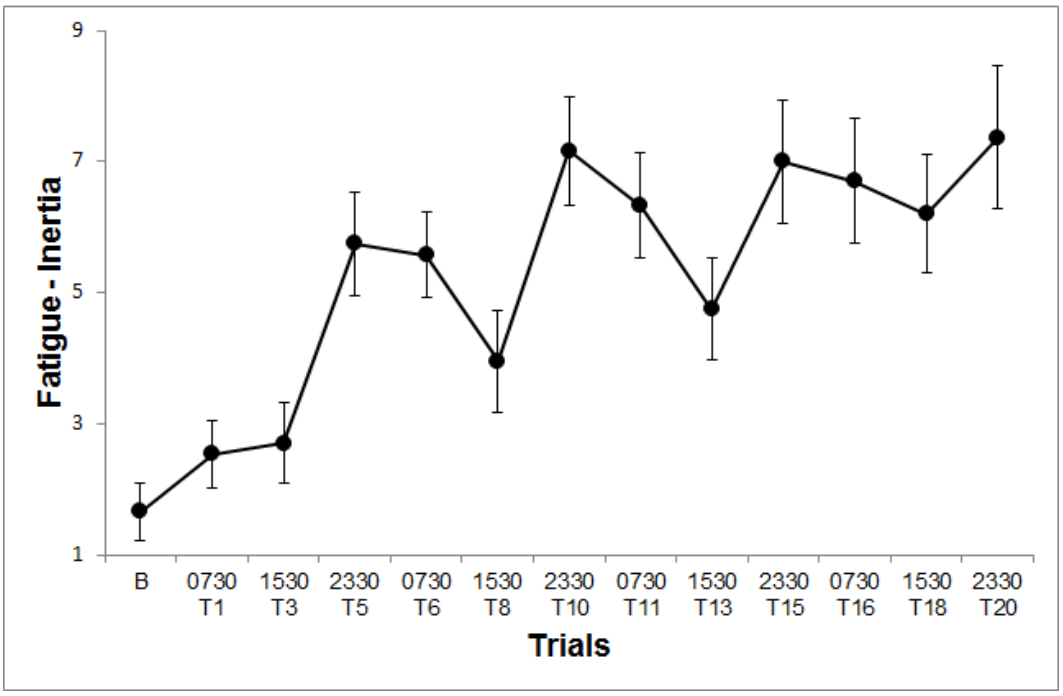


Figure 27. POMS Fatigue-Inertia scores across the Experimental Phase. Post-hoc analyses revealed that there was a significant difference between participants' responses during baseline and their responses to Trials 3 – 20.

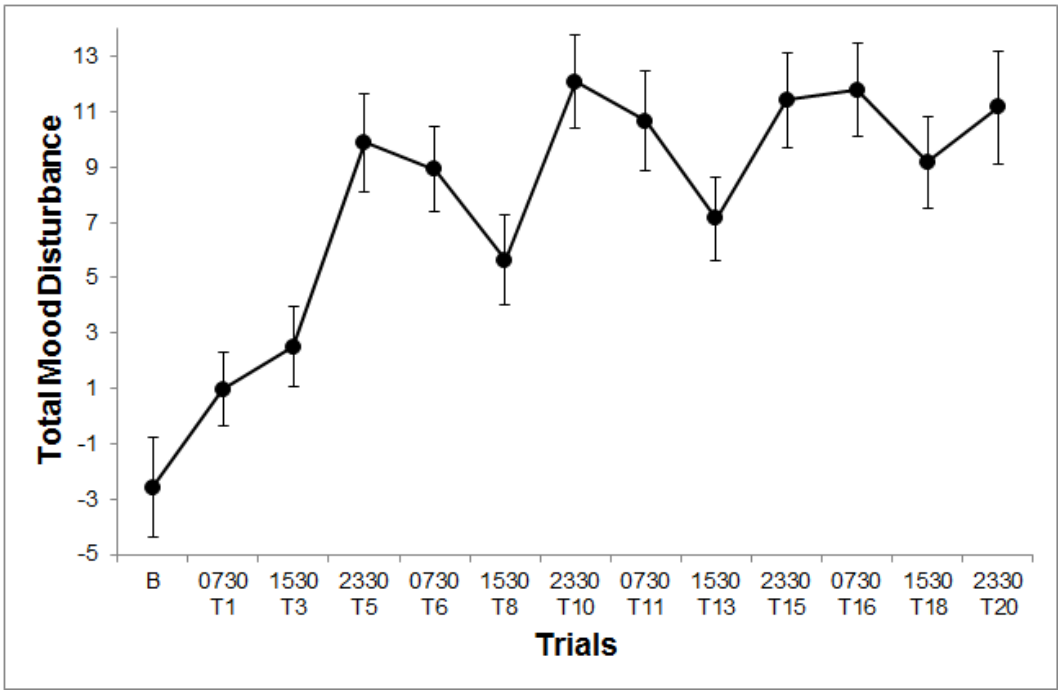


Figure 28. POMS Total Mood Disturbance scores across the Experimental Phase. Post-hoc analyses revealed that there was a significant difference between participants' responses during baseline and their responses on each of the trials during the sleep restriction period.

*Fatigue Avoidance Scheduling Tool (FAST).* Participants' predicted performance effectiveness changed significantly over the course of the study, as shown below in Table 11 and Figure 29. Post-hoc analyses indicated that predicted effectiveness on Trial 1 was significantly different than that predicted for Trials 2 – 8 and 10 – 20. During each day of the sleep restriction period, the lowest predicted level of performance effectiveness was observed at 2330, the last test session of the day. Thus FAST recognized that participants' effectiveness would be lowest during the 2330 test session and would gradually decrease across the sleep restriction period.

*Table 11*. ANOVA results for the Fatigue Avoidance Scheduling Tool

| F | df | *p* | $\eta_p^2$ |
|---|---|---|---|
| 101.167 | (19, 437) | 0.000* | 0.815 |

*Significant at the .05 level

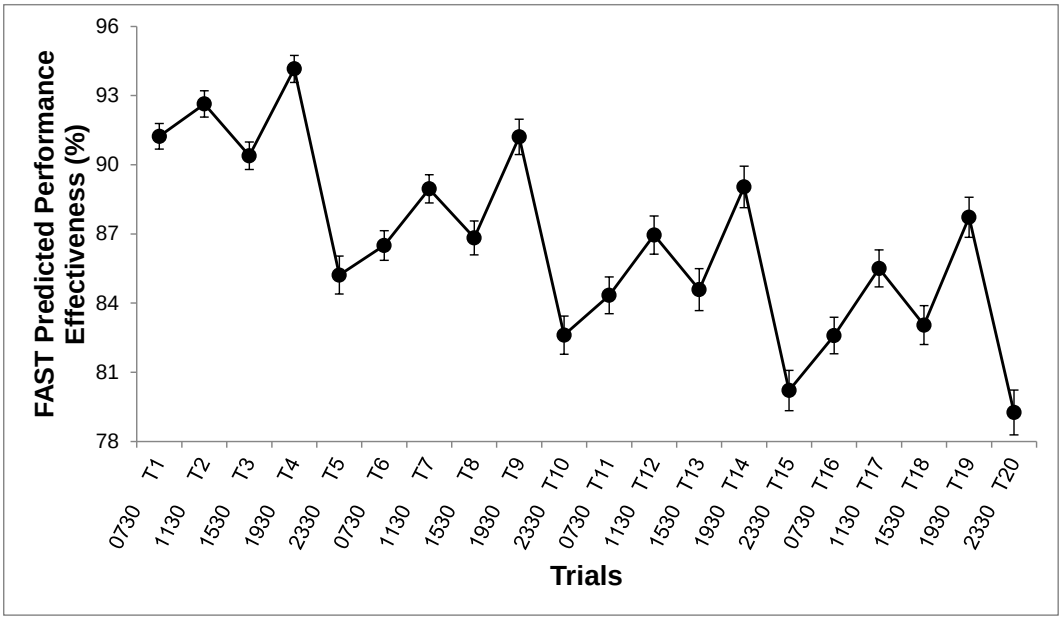


Figure 29. Change in participants' predicted effectiveness over the course of the study. Predicted performance at Trial 1 was significantly different than for Trials 2 – 8 and 10 – 20.

*University of Pennsylvania Smell Identification Test (UPSIT).* Results for the UPSIT measure indicated that participants' ability to identify the 40 test scents was not impaired (Figure 30). Specifically, participants actually demonstrated improved ability to identify the scents from the pre-test (M = 33.83, SD = 4.30) to the post-test (M = 34.08, SD = 3.41) which was completed after 4 days of sleep restriction, t(23) = -.319, $p$ = .752 .

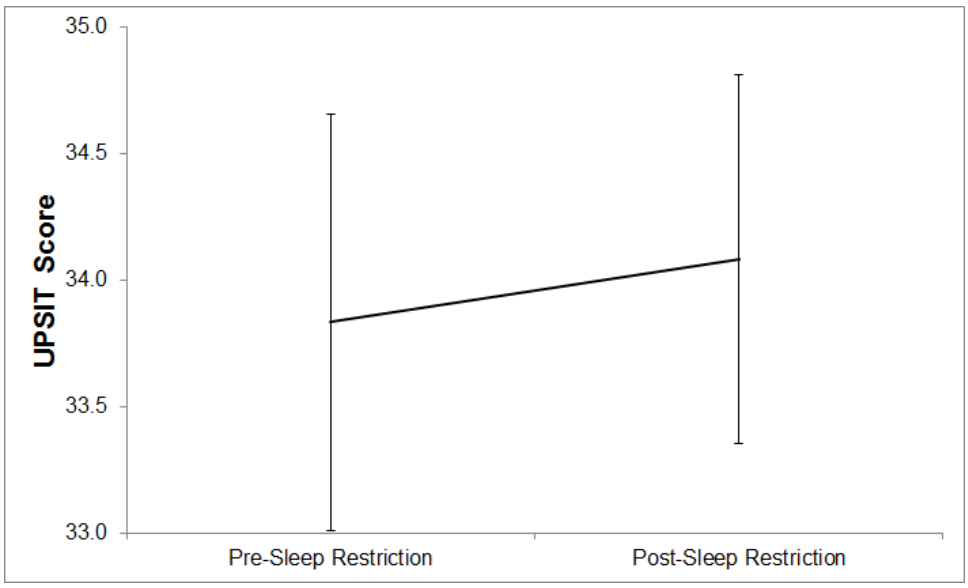


Figure 30. Change in participants' score on the University of Pennsylvania Smell Identification Test (UPSIT) from the baseline test to the final test before recovery sleep.

***Stage 1 Summary.*** During Stage 1, a series of repeated measures ANOVAs were conducted to detect any gradual increases in fatigue or performance impairment throughout the course of the study. Despite repeated training sessions, several of the tasks indicated practice effects such that participants' response time (e.g., Shifting Accuracy and Divided Attention measures of the Flight Fit test) and recall accuracy (Flight Fit Short Term Memory task and the dual *n*-back test) improved significantly over the course of the Experimental Phase. Results such as these suggest that some cognitive performance measures are not sensitive to the effects of fatigue due to chronic sleep restriction. Conversely, effects on many of the other measures replicated existing literature by revealing a gradual deterioration in performance over the course of the sleep loss period. These changes included significant deviations from baseline on the subcomponents of the PMI FIT 2000 and some subscales of the POMS, as well as increased subjective sleepiness as measured by the SSS. Further, FAST performance estimates closely followed the pattern of performance deterioration observed on tasks such as the PVT and flight simulator across the sleep restriction period.

Additionally, comparison of the figures suggests that some of these measures were more sensitive to fatigue due to circadian influences whereas others appeared to be more sensitive to participants' increasing sleep debt. For example, Pupil Constriction Amplitude (Figures 6) revealed a significant response to troughs in the circadian cycle. Conversely, other measures, such as Pupil Diameter (Figures 7) and total lapse time on the flight simulator (Figure 10) appeared to be more influenced by the homeostatic sleep drive, which is the pressure to sleep which gradually increases the longer someone has been awake. Specifically, data from Pupil Constriction Amplitude revealed peaks for the evening testing sessions, whereas time-off-target on the flight simulator exhibited more peaks during the early morning test session, though both measures had increasing deviations from baseline which indicated worsening fatigue effects over the course of the Experimental Phase. Further, subjective response measures such as the Stanford Sleepiness Scale (Figure 21) and several components of the POMS (Figures 26 – 28) appeared to effectively track the effects of both the circadian and homeostatic processes over the

period of sleep restriction, with spikes in highly fatigue-related behavior during circadian troughs, as well as growing deviation from baseline indicating a general decline due to homeostatic processes. However, these results should be interpreted with a measure of caution as these statements to which participants responded are subject to individual interpretation as well as intentional misreporting or deception.

Taken as a whole, these findings indicate that no single tool should be expected to accurately evaluate an individual's level of fatigue or readiness for duty. Cognitive, subjective, and physiological assessments ought to be included to determine whether someone is too fatigued to perform adequately. Further, as demonstrated above, different measures are sensitive to different cognitive processes, so while an individual may not appear to have any impairment on one measure, another task may reveal impairment for a different cognitive function, indicating that they should not be expected to perform their role in a mission safely. It is also vital that any effective readiness for duty tool be sensitive to fluctuations in both circadian and homeostatic processes.

The results of the Stage 1 analyses revealed several cognitive and physiologic decrements when fatigued due to sleep restriction. Though this information is invaluable in assessing the severity of the fatigue experienced by participants and emphasizing the importance of proper sleep schedules, it does little to demonstrate any individual differences in fatigue susceptibility. In other words, the results from Stage 1 tell us only that participants on average experienced impairment due to the sleep restriction, and consequently mask any individuals who were more or less adversely affected. To address this limitation of the ANOVA, a form of regression modeling was used in Stage 2 to identify differences at the individual level using the measures which were influenced by fatigue in Stage 1.

**Stage 2**

Stage 2 of the analyses addresses the matter of determining which measures are sensitive to individual differences in fatigue susceptibility in response to chronic sleep restriction. Bivariate Hierarchical Linear Modeling (HLM), a form of regression modeling, was utilized for this phase of data analysis. With this type of modeling, it is possible to "nest" data within participants to better identify individual differences. Measures analyzed at this stage included total lapse time on the flight simulator, Stanford Sleepiness Scale (SSS) responses, predicted effectiveness from FAST, as well as components of the PMI FIT 2000 and Profile of Mood States (POMS). The results of these analyses were used to develop the subsequent algorithms in Stage 3.

***Bivariate Hierarchical Linear Models***. Hierarchical Linear Modeling (HLM), also known as multilevel modeling, is useful in identifying common traits among different individuals. The outcome of these analyses reveals how effectively each of the predictor variables (e.g., SSS, FAST, lapse time on the flight simulator, etc.) is able to predict the changes in the criterion variable, the average number of PVT lapses. The number of lapses in attention on the PVT was selected as the criterion variable because this measure is recognized as a valid and highly reliable objective assessment of fatigue. Predictor variables for these analyses were any measures from Stage 1 which were significantly influenced by fatigue.

All variables that showed signs of a significant bivariate relation at Level 1 or Level 2 with mean number of PVT lapses are identified in Tables 12 and 13, respectively. In a bivariate HLM, estimates of commonality are made between each of the predictor variables and the criterion variable on two levels. Significance for Level 1 (fixed effects) for one of the predictor variables would indicate that when the data from all participants are collapsed into a group, that predictor variable accurately predicts the criterion variable. For the present analyses, significant results for Level 1 analyses would mean that the selected predictor variable can be used to effectively predict changes in fatigue, as assessed by the PVT, during 4 days of sleep restriction. Further, significance for Level 2 (random effects) would indicate that there were significant individual differences in the predictions for the criterion variable. In other words, significance for a given predictor at Level 2 would suggest that this predictor was sensitive to individual differences in susceptibility to the effects of fatigue in a chronic sleep restriction situation. If no significant inter-individual variability was found (non-significance at Level 2), the predictor variable was re-analyzed excluding the random effects and the outcome of this analysis was reported for Level 1.

*Table 12*. Level 1 Bivariate HLMs Relation with PVT Lapses as the Outcome

| | Equation | t | df | *p* |
|---|---|---|---|---|
| PMI Saccadic Velocity | Y = P0 + P1 * (PMI_SV) + R | -2.461 | 23 | .022* |
| PMI Pupil Diameter | Y = P0 + P1 * (PMI_DIAM) + R | -1.304 | 23 | .205 |
| PMI Constriction Latency⊗ | Y = P0 + P1 * (PMI_LAT) + R | .837 | 23 | .403 |
| PMI Constriction Amplitude⊗ | Y = P0 + P1 * (PMI_AMP) + R | 1.159 | 23 | .247 |
| Flight Sim Total Lapse Time | Y = P0 + P1 * (FS_TOTAL) + R | 3.858 | 23 | .001* |
| Stanford Sleepiness Scale⊗ | Y = P0 + P1 * (SSS) + R | 6.136 | 23 | .000* |
| POMS Vigor/Anxiety⊗ | Y = P0 + P1 * (POMS_V) + R | -5.227 | 23 | .000* |
| POMS Fatigue/Inertia | Y = P0 + P1 * (POMS_F) + R | 3.056 | 23 | .006* |
| POMS Total Mood Disturbance | Y = P0 + P1 * (POMS_TMD) + R | 3.555 | 23 | .002* |
| FAST | Y = P0 + P1 * (FAST) + R | -4.317 | 23 | .000 |

***Note:*** PMI_SV = Saccadic Velocity, PMI_DIAM = Diameter, PMI_AMP = Constriction Amplitude, PMI_LAT = Constriction Latency, FS_TOTAL = Flight Simulator Total Lapse Time, SSS = Stanford Sleepiness Scale, POMS_V-A = POMS Vigor-Activity, POMS_F-I = POMS Fatigue-Inertia, POMS_TMD = POMS Total Mood Disturbance, FAST = Fatigue Avoidance Scheduling Tool

*Significant at the .05 level

⊗Results are from analyses which excluded random effects

*Table 13*. Level 2 Bivariate HLMs Relation with PVT Lapses as the Outcome

| | Equation | $\chi 2$ | df | *p* |
|---|---|---|---|---|
| PMI Saccadic Velocity | P0 = B00 + R0<br>P1 = B10 + R1 | 45.076 | 23 | .004* |
| PMI Pupil Diameter | P0 = B00 + R0<br>P1 = B10 + R1 | 74.931 | 23 | .000* |
| PMI Constriction Latency | P0 = B00 + R0<br>P1 = B10 + R1 | 24.800 | 23 | .360 |
| PMI Constriction Amplitude | P0 = B00 + R0<br>P1 = B10 + R1 | 24.938 | 23 | .353 |
| Flight Sim Total Lapse Time | P0 = B00 + R0<br>P1 = B10 + R1 | 161.818 | 23 | .000* |
| Stanford Sleepiness Scale | P0 = B00 + R0<br>P1 = B10 + R1 | 20.315 | 23 | > .500 |
| POMS Vigor/Anxiety | P0 = B00 + R0<br>P1 = B10 + R1 | 32.378 | 23 | .071 |
| POMS Fatigue/Inertia | P0 = B00 + R0<br>P1 = B10 + R1 | 55.971 | 23 | .000* |
| POMS Total Mood Disturbance | P0 = B00 + R0<br>P1 = B10 + R1 | 53.418 | 23 | .001* |
| FAST | P0 = B00 + R0<br>P1 = B10 + R1 | 59.898 | 23 | .000* |

*Significant at the .05 level

*PMI Saccadic Velocity*. The equations for both Level 1 and Level 2 were significant for the PMI FIT component Saccadic Velocity. Overall, participants who had the least variability in their number of PVT lapses also tended to have the least amount of variability in Saccadic Velocity as measured by the PMI (Figure 31), and this trend was most evident among those participants who had higher numbers of PVT lapses.

*PMI Pupil Diameter*. When included as a predictor variable, the PMI FIT component Pupil Diameter was not significant for the Level 1 analysis, but the results were significant for the Level 2 analyses. Participants who had fewer lapses on the PVT, as well as less variability in their performance, also exhibited less change in pupil diameter (Figure 32) compared to participants who had greater variability in their PVT performance.

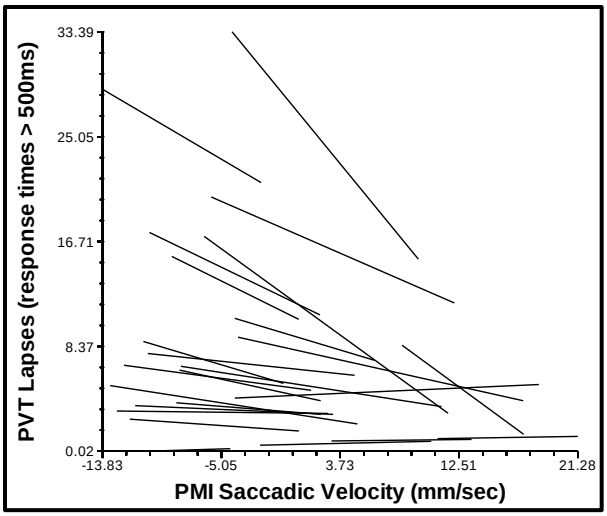


Figure 31. Individual slopes for PVT Lapses in relation to Saccadic Velocity measured by the PMI in millimeters per second (group mean-centered values). There was a significant group effect and significant individual differences, such that, lapses tended to increase as Saccadic Velocity increased though there were significant individual differences in this trend.

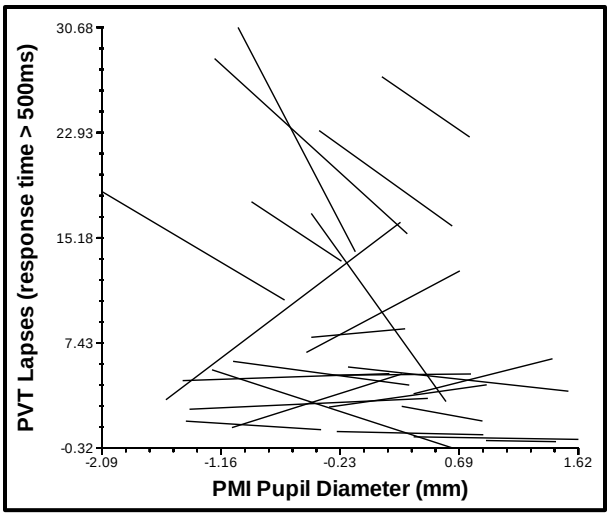


Figure 32. Individual slopes for PVT lapses in relation to responses to Pupil Diameter measured by the PMI in millimeters (group mean-centered values). There was no significant group effect, but there were significant individual differences, which showed, on average, participants who had little variance in diameter also exhibited fewer PVT lapses and less variance in their performance.

*PMI Constriction Latency*. As observed with the Pupil Diameter analyses, the results for the PMI measure Constriction Latency were not significant for Level 1 but were significant for the Level 2 model. These results are shown in Figure 33 and, in general, suggest that participants who had fewer lapses in attention on the PVT also had a shorter latency period for pupil constriction when compared to participants who had a greater number of lapses.

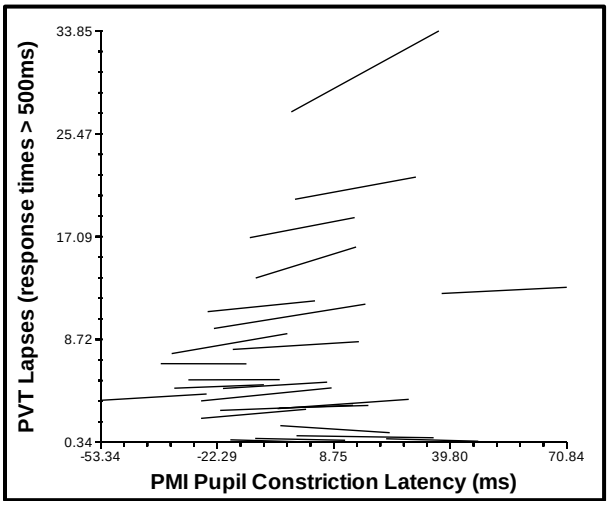


Figure 33. Individual slopes for PVT lapses in relation to Pupil Constriction Latency measured by the PMI in milliseconds (group mean-centered values). No differences were evident at the group level, but there were significant individual differences between PVT lapses and pupil constriction latency. Specifically, participants with shorter pupil constriction latencies tended to have fewer PVT lapses than participants with longer constriction latencies.

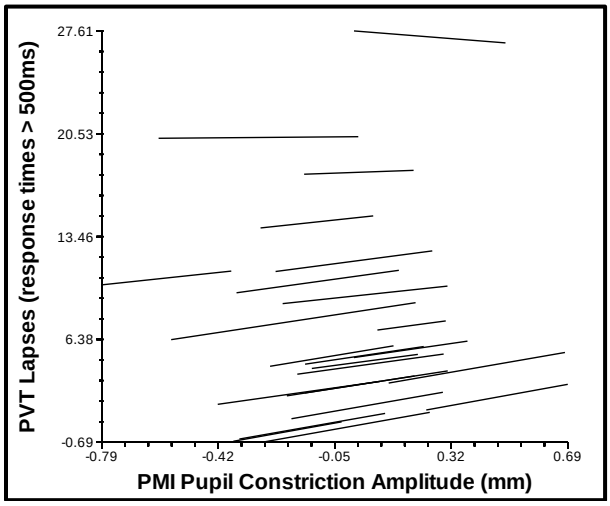


Figure 34. Individual slopes for PVT lapses in relation to Pupil Constriction Amplitude in millimeters (group mean-centered values). There was no significant group effect but there were significant individual differences, such that participants with greater variability in their pupil constriction amplitude tended to have greater variability in their PVT lapses.

*PMI Constriction Amplitude*. The last of the four PMI component measures, Constriction Amplitude, was also not a significant predictor of PVT lapses at the group level during 4 days of chronic sleep restriction. However, results for the analysis at the individual level were significant and are shown in Figure 34. As a whole, participants who demonstrated greater variability in the number of PVT lapses also demonstrated greater variability in constriction amplitude.

*Flight Simulator Total Lapse Time*. The equation for Level 1 revealed that the total amount of time that participants' attention lapsed during the flight simulation could accurately predict the number of lapses on the PVT. Specifically, the total number of PVT lapses increased linearly as the number of lapses in the flight simulator increased. Likewise, the equation for Level 2 was also significant. Inspection of the inter-individual slope variability (Figure 35) indicates that, like Saccadic Velocity, those participants who demonstrated the least amount of variability in the number of lapses during the flight simulation also demonstrated the least amount of variability in the number of PVT lapses.

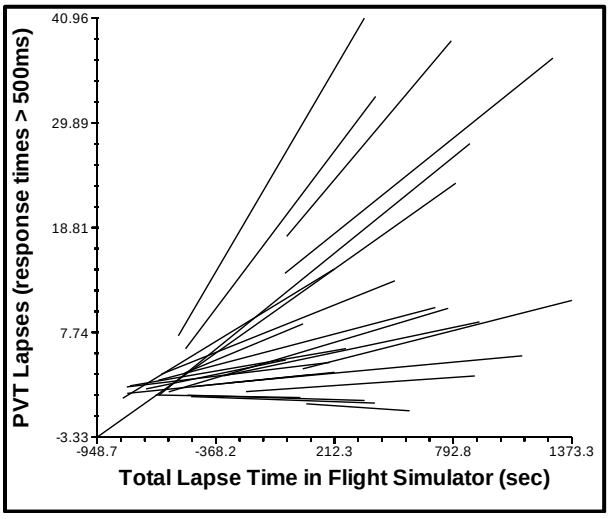


Figure 35. Individual slopes for PVT lapses in relation to the total time lapse during flight simulation in seconds. There was a significant group effect as well as significant individual differences such that, on average, as total lapse time in the flight simulator increased, the number of PVT lapses also increased.

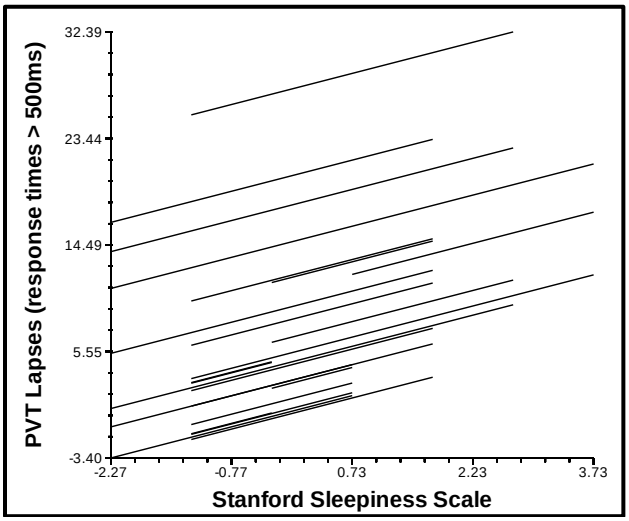


Figure 36. Individual slopes for PVT lapses in relation to responses to the Stanford Sleepiness Scale (group mean-centered values). There was a significant group effect but there were no significant individual differences between the individual slopes, indicating that the relationship between PVT lapses and responses to the Stanford Sleepiness Scale were similar for all participants.

*Stanford Sleepiness Scale (SSS)*. The fixed effects for the Stanford Sleepiness Scale were significant, though the same was not true for the random effects (see Figure 36). In other words, although participants' responses to the SSS did change significantly as the number of PVT lapses changed, there was no significant difference between the individual slopes. Since the Level 2 equation using random effects was not significant, the value reported for Level 1 utilized a re-estimation of the model without the random effects. The lack of a significant random effects suggest that increases in sleepiness as measured using the SSS in relation to PVT lapses is best conceptualized at the group level.

*POMS Vigor-Activity*. The Level 1 equation for the Profile of Mood States component Vigor-Activity was significant, indicating that subjective assessments of energy levels may be predictive of lapses in attention at the group level (Figure 37). However, analyses revealed that

the Level 2 equation was not significant, and thus the results reported from Level 1 are taken from a re-estimation of the model excluding random effects. The absence of significant individual differences in the predictive value of the POMS Vigor-Activity component suggests that most people may be fairly equally matched in terms of the extent to which subjective feelings relate to actual performance.

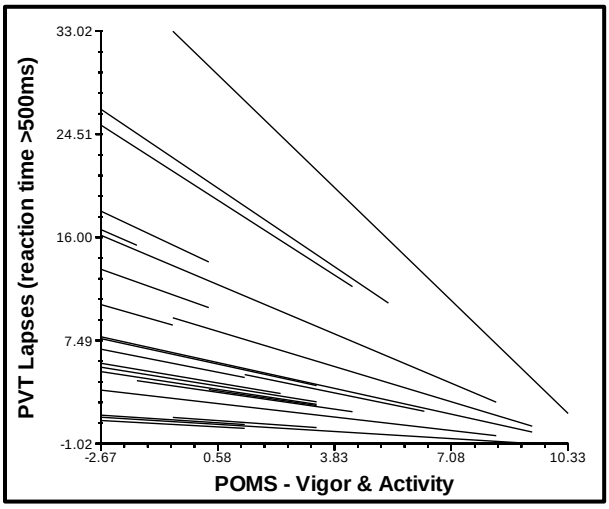


Figure 37. Individual slopes of PVT lapses in relation to Vigor-Activity measured by the POMS (group mean-centered values). There was a significant group effect, with vigor decreasing as the number of lapses on the PVT increased. However, there were no significant individual differences.

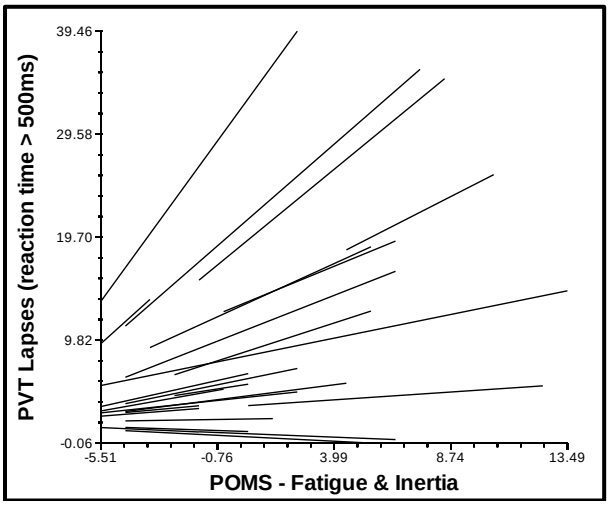


Figure 38. Individual slope of PVT lapses in relation to Fatigue and Inertia measured using the POMS (group mean-centered values). There was a significant group effect, as well as significant individual differences, such that, the greater the variability in the performance on the Fatigue and Inertia portion of the POMS, the greater the number of lapses on the PVT.

*POMS Fatigue-Inertia.* For the Fatigue-Inertia component of the POMS, the model was significant for both Level 1 and Level 2. At the group level, higher self-ratings for this measure of sleepiness and lethargy were related to a greater number of lapses in attention on the PVT. Additionally, the results of the Level 2 analysis indicated that there were significant individual differences in response to chronic sleep restriction. As evident in Figure 38, participants who had a higher number of lapses on the PVT also tended to have greater variability in their performance on this task. Further, these participants also tended to report higher levels of sleepiness than did participants who had fewer lapses in attention.

*POMS Total Mood Disturbance*. As the Total Mood Disturbance is a composite score reflecting the six subscales of the POMS, it is not surprising that the predictive value of this measure for the number of PVT lapses was significant at both the group and individual level. In general, higher Total Mood Disturbance scores appear to be related to a greater number of lapses in attention (Figure 39). Additionally, results from the Level 2 equation indicate that, as observed on the Fatigue-Inertia subscale, participants who had a higher number of lapses also demonstrated greater variability in their performance while reporting greater mood disturbance.

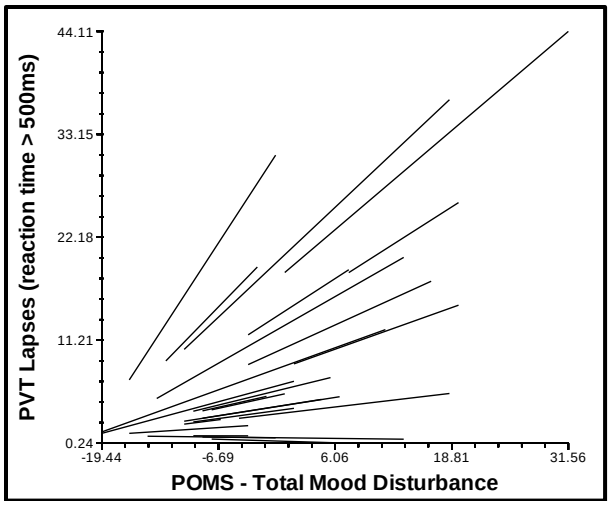


Figure 39. Individual slopes for PVT lapses in relation to Total Mood Disturbance measured by the POMS (group mean-centered values). There was a significant group effect as well as significant individual differences, such that, on average, the higher the Total Mood Disturbance score, the greater the number of PVT lapses.

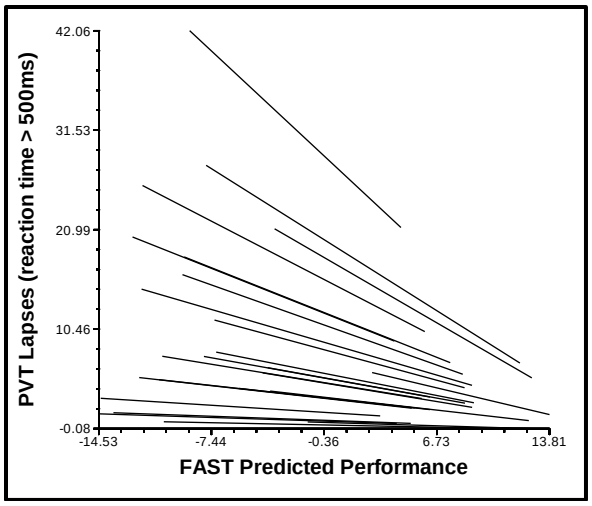


Figure 40. Individual slopes for PVT lapses in relation to predicted performances in FAST (group mean-centered values). There was a significant group effect as well as significant individual differences. On average, participants who were predicted to have the greatest performance impairment according to FAST demonstrated the greatest number of PVT lapses.

*Fatigue Avoidance Scheduling Tool (FAST)*. Both Level 1 and Level 2 equations using the FAST were able to significantly predict the number of PVT lapses. Specifically, the results for the Level 1 equation revealed that the deterioration in performance predicted by FAST corresponded to an increase in lapses as measured by the PVT. Likewise, results for the Level 2 equation revealed significant individual differences such that those participants who were predicted to have the least performance deficit according to FAST demonstrated the fewest number of lapses on the PVT (Figure 40).

***Stage 2 Summary***. Taken as a whole, the results from this series of analyses suggested that several of the subjective and objective measures included in this study may be useful in predicting lapses in attention at both the group and individual level during a period of chronic sleep restriction. The significant subjective factors were two measures obtained from the POMS, Fatigue-Inertia and Total Mood Disturbance. For these factors, greater fatigue or mood disturbance tended to predict a greater number of lapses on the PVT as well as increased variability in performance. Because the Stanford Sleepiness Scale (SSS) is specifically designed to assess subjective fatigue levels, it is interesting that this measure was not a significant predictor of PVT lapses at the individual level. One explanation for this may simply be the nature of the different measures. Specifically, the Fatigue-Inertia score is based on the participants' response to seven adjectives and Total Mood Disturbance is based on responses to 65 different adjectives, but the SSS asks for participants to respond to a single statement. This is not to suggest that the SSS is a poor measure of fatigue but simply that it may not be possible for a response to one statement to capture the variability in individual differences in response to chronic sleep restriction.

The remaining measures which were significant predictors of PVT lapses at both the group and individual level were objective, and included one measure of performance and one physiologic assessment. The performance measure was the total lapse time on the X-Plane flight simulator, which of all the neurobehavioral assessments participants in this study completed, was

the only measure that was sensitive to the performance deterioration over the chronic sleep restriction period. Finding that performance on this measure is strongly predictive of the total number of lapses on the PVT supports the use of a flight simulator as a more operationally-relevant means of detecting fatigue. These results suggest that lapse time on a flight simulator is not subject to the same practice effects which seemed to influence performance on other measures, such as those taken from the Flight Fit test battery, and thus may be a valuable tool in developing a readiness-to-perform assessment.

The best objective, physiologic measure was part of the PMI oculometric assessment, Saccadic Velocity. The PMI FIT typically takes less than a minute to complete and can be done with little to no oversight or supervision from research staff. This finding replicates existing research concluding that sleep loss leads to reduced saccadic velocity, though the degree of this decrease in speed differs between individuals (Chandler, Arnold, Phillips, Lojewski, & Horning, 2010; Rowland et al., 2005). Conclusions from this previous research as well as the findings of the present study suggest that oculometric measures, such as those recorded by the PMI, might be time-efficient means of identifying individuals who are too fatigued to perform at the best of their abilities. Moreover, combining these measures with other factors, such as lapse time on the flight simulator, would greatly strengthen their predictive efficacy.

The final two predictive factors were both related to participants' sleep and wake patterns over the course of the study – time (i.e., test session) and FAST-predicted effectiveness score. Significance at both the group and individual levels for the first of these measures, time, confirmed that the gradually increasing sleep debt participants experienced during the study was related to a growing performance impairment, but that there were strong individual differences in the severity of this impairment. The results were similar for the FAST scores in that predictions of poorer performance corresponded to increased number of lapses on the PVT. However, as noted previously, there were strong individual differences with some participants showing little to no change in performance whereas the performance of others deteriorated drastically. These results suggest that a critical part of predicting performance during periods of insufficient sleep may be having a comprehensive picture of the individual's sleep/wake patterns.

**Stage 3**

Having identified which of the included factors had the greatest predictive value for performance impairments over 4 days of chronic sleep restriction, the final stage of these analyses was intended to develop predictive algorithms. Specifically, five of the factors which were able to predict both group and individual changes in the number of lapses on the PVT were further examined through a series of enter-method linear regression analyses to determine whether combining these factors might improve their predictive value. The first set of analyses included participants across all test sessions, whereas subsequent analyses constrained the data based on different factors, such as Extraversion facet scores on the NEO-PI-R and testing session, to determine whether selected measures might have greater predictive value under different conditions. For all of these analyses, a regression was first done using data from the FAST performance estimates, followed by an analysis including all five of the predictive factors. If the model including all five factors explained less than 20% (i.e., $R^2 \leq .20$) of the variance, no further analyses were conducted. If the five factors explained more than 20% of the variance, additional analyses were done to examine the predictive value of algorithms including the three

factors with the greatest beta coefficients. In addition to the models described below, other analyses were conducted which examined the efficacy of models constrained by the other NEO-PI-R scores, but none of these yielded models which satisfied the standard of 20% variance explained.

When all of the data were analyzed together, the model created using just the FAST performance estimates was significant, but it predicted little more than 1% of the variance (Table 14). The model was strengthened with the addition of the four other predictors – total lapse time on the flight simulator, PMI Saccadic Velocity, and POMS factors Fatigue/Inertia and Total Mood Disturbance – with total variance explained increasing to almost 12%. This represents a clear improvement over using the FAST scores alone to predict impairments during chronic sleep restriction, but still offers very little towards predicting performance decrements.

Table 14. Linear Regression with FAST Performance Estimate as Primary Predictor of PVT Lapses.

| | **Equation** | $\mathbf{R^2}$ | **Δ F** | ***df* 1** | ***df* 2** | ***p*** |
|---|---|---|---|---|---|---|
| FAST | PVT_lapses = FAST * -.104 | 0.011 | 5.180 | 1 | 478 | .023 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .045) + (FS_TotalLapse * .183) + (PMI_SV * .005) + (POMS_F/I * -.071) + (POMS_TMD * .313) | 0.117 | 7.448 | 5 | 282 | .000 |

Note: FAST = Fatiuge Avoidance Scheduling Tool; FS_TotalLapse = Flight Simulator Total Lapse Time; PMI_SV = PMI Saccadic Velocity; POMS_F/I = POMS Fatigue/Inertia; POMS_TMD = POMS Total Mood Disturbance

The second series of analyses was conducted in three phases, first analyzing data from those participants who were ranked in the lowest 25% on the NEO-PI-R Gregariousness facet (E2) of the Extraversion trait, followed by those ranked in the middle 50%, and finally those in the highest 25% (Table 15). For all three groups, the results of the model which included only FAST were somewhat better than was observed previously, with the performance estimate predicting 1.5 – 4.1% of the variance. Among participants in the middle 50% group, results including all five of the predictors were also similar, with this model still predicting less than 5% of the variance. However, the most interesting models were found for the two extreme groups, those in the lowest and highest 25%. For these groups, the model built using all five predictors accounted for nearly 68% and nearly 50% of the total variance, respectively. Subsequent analyses determined that for participants who were ranked low on the Gregariousness facet, total lapse time on the flight simulator, PMI Saccadic Velocity, and Total Mood Disturbance on the POMS were the three most predictive factors. The best three-predictor model for those ranking high on the Gregariousness facet also included total lapse time on the flight simulator and Total Mood Disturbance on the POMS, as well as the Fatigue/Inertia measure from the POMS.

Table 15. Linear Regression with FAST Performance Estimate as Primary Predictor of PVT Lapses and Data Constrained by Gregariousness Facet (E2) of Extraversion Trait.

| | Equation | $R^2$ | Δ F | *df* 1 | *df* 2 | *p* |
|---|---|---|---|---|---|---|
| Participants Who Scored in the Lowest 25% on the Gregariousness Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * -.190 | 0.036 | 4.427 | 1 | 118 | .083 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .109) + (FS_TotalLapse * .400) + (PMI_SV * -.126) + (POMS_F/I * -.196) + (POMS_TMD * .880) | 0.679 | 27.886 | 5 | 66 | .000 |
| FS_TotalLapse, PMI_SV, POMS_TMD | PVT_lapses = (FS_TotalLapse * .384) + (PMI_SV * -.068) + (POMS_TMD * .633) | 0.653 | 42.746 | 3 | 68 | .000 |
| Participants Who Scored in the Middle 50% on the Gregariousness Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * -.202 | 0.041 | 10.115 | 1 | 238 | .000 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * -.151) + (FS_TotalLapse * -.073) + (PMI_SV * -.124) + (POMS_F/I * .121) + (POMS_TMD * -.075) | 0.044 | 1.282 | 5 | 138 | .275 |
| Participants Who Scored in the Higest 25% on the Gregariousness Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * .124 | 0.015 | 1.835 | 1 | 118 | .178 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .054) + (FS_TotalLapse * .622) + (PMI_SV * .021) + (POMS_F/I * -.501) + (POMS_TMD * .337) | 0.495 | 12.941 | 5 | 66 | .000 |
| FS_TotalLapse, POMS_F/I, POMS_TMD | PVT_lapses = (FS_TotalLapse * .638) + (POMS_F/I * -.499) + (POMS_TMD * .310) | 0.493 | 21.999 | 3 | 68 | .000 |

Note: FAST = Fatiuge Avoidance Scheduling Tool; FS_TotalLapse = Flight Simulator Total Lapse Time; PMI_SV = PMI Saccadic Velocity; POMS_F/I = POMS Fatigue/Inertia; POMS_TMD = POMS Total Mood Disturbance

Also examining a subcomponent of the Extraversion trait from the NEO-PI-R, the next series of analyses tested whether an individual's rank on the Activity facet (E4) could predict their fatigue susceptibility to chronic sleep restriction. As was observed with the Gregariousness facet (Table 16), Activity had the greatest predictive value for individuals in one of the two extreme categories rather than for participants in the middle 50% on the measure. For the middle 50%, FAST performance estimates predicted less than 3% of the variance in the number of PVT lapses, while adding in the other four factors improved this to 11.8%. Conversely, for participants categorized in the lowest 25% of the Activity facet, FAST predicted a meaningful amount of the variance, a little more than 9%. Additionally, including all five factors into the algorithm increased the predictive value to 37.3%. Of the five factors, the most predictive for this group were PMI Saccadic Velocity and both the POMS Fatigue / Inertia and Total Mood Disturbance factors. Although FAST alone had a lower predictive value when data was constrained to just participants in the higher 25% group (0.8%), the algorithm including all five factors was much more effective, predicting nearly 28% of the variance. As observed with

participants ranked in the lowest 25%, for those in the highest 25% group according to their Activity facet rating, the three best predicting factors were Saccadic Velocity, POMS Fatigue / Inertia, and POMS Total Mood Disturbance.

Table 16. Linear Regression with FAST Performance Estimate as Primary Predictor of PVT Lapses and Data Constrained by Activity Facet of Extraversion Trait.

| | Equation | $R^2$ | Δ F | *df* 1 | *df* 2 | *p* |
|---|---|---|---|---|---|---|
| Participants Who Scored in the Lowest 25% on the Activity Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * -.303 | 0.092 | 11.895 | 1 | 118 | .001 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * -.024) + (FS_TotalLapse * .017) + (PMI_SV * -.233) + (POMS_F/I * -.053) + (POMS_TMD * .666) | 0.373 | 7.840 | 5 | 66 | .000 |
| PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (PMI_SV * -.243) + (POMS_F/I * -.042) + (POMS_TMD * .633) | 0.372 | 13.417 | 3 | 68 | .000 |
| Participants Who Scored in the Middle 50% on the Activity Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * -.154 | 0.024 | 5.804 | 1 | 238 | .017 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .012) + (FS_TotalLapse * .199) + (PMI_SV * -.010) + (POMS_F/I * .098) + (POMS_TMD * .144) | 0.118 | 3.695 | 5 | 138 | .004 |
| Participants Who Scored in the Higest 25% on the Activity Facet of the Extraversion Trait | | | | | | |
| FAST | PVT_lapses = FAST * .087 | 0.008 | 0.892 | 1 | 118 | .347 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .084) + (FS_TotalLapse * .204) + (PMI_SV * .287) + (POMS_F/I * -.824) + (POMS_TMD * .782) | 0.275 | 5.019 | 5 | 66 | .001 |
| PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (PMI_SV * .302) + (POMS_F/I * -1.053) + (POMS_TMD * 1.007) | 0.242 | 7.243 | 3 | 68 | .000 |

Note: FAST = Fatiuge Avoidance Scheduling Tool; FS_TotalLapse = Flight Simulator Total Lapse Time; PMI_SV = PMI Saccadic Velocity; POMS_F/I = POMS Fatigue/Inertia; POMS_TMD = POMS Total Mood Disturbance

The final series of predictive analyses constrained data based on test session, using only the testing times during which the POMS was administered – the first (0730), third (1530), and fifth (2330) sessions each day. As evident in Table 17 below, the models created from FAST data collected during the 1530 and 2330 testing sessions failed to predict a meaningful portion of the variance, though the models were somewhat improved with the addition of the other four predictors. However, the results from data recorded during the earliest test sessions were much more promising, with FAST alone explaining almost 5% of the variance and the five combined predictors explaining 22.4% of the variance. Further analyses confirmed that the three most effective predictors in the model were total lapse time on the flight simulator, PMI Saccadic Velocity, and POMS Total Mood Disturbance.

Table 17. Linear Regression with FAST Performance Estimate as Primary Predictor of PVT Lapses and Data Constrained by Testing Session.

| | Equation | $R^2$ | Δ F | *df* 1 | *df* 2 | *p* |
|---|---|---|---|---|---|---|
| | Test Sessions Beginning at 0730 | | | | | |
| FAST | PVT_lapses = FAST * .217 | 0.047 | 4.649 | 1 | 94 | .034 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * -.020) + (FS_TotalLapse * .293) + (PMI_SV * -.023) + (POMS_F/I * .017) + (POMS_TMD * .271) | 0.224 | 5.207 | 5 | 90 | .000 |
| FS_TotalLapse, PMI_SV, POMS_TMD | PVT_lapses = (FS_TotalLapse * .300) + (PMI_SV * -.019) + (POMS_TMD * .290) | 0.224 | 8.846 | 3 | 92 | .000 |
| | Test Sessions Beginning at 1530 | | | | | |
| FAST | PVT_lapses = FAST * -.034 | 0.001 | 0.110 | 1 | 94 | .741 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * .083) + (FS_TotalLapse * -.043) + (PMI_SV * -.028) + (POMS_F/I * -.125) + (POMS_TMD * .330) | 0.165 | 3.545 | 5 | 90 | .006 |
| | Test Sessions Beginning at 2330 | | | | | |
| FAST | PVT_lapses = FAST * .015 | 0.000 | 0.020 | 1 | 94 | .887 |
| FAST, FS_TotalLapse, PMI_SV, POMS_F/I, POMS_TMD | PVT_lapses = (FAST * -.010) + (FS_TotalLapse * .109) + (PMI_SV * .090) + (POMS_F/I * -.325) + (POMS_TMD * .396) | 0.097 | 1.929 | 5 | 90 | .097 |

Note: FAST = Fatiuge Avoidance Scheduling Tool; FS_TotalLapse = Flight Simulator Total Lapse Time; PMI_SV = PMI Saccadic Velocity; POMS_F/I = POMS Fatigue/Inertia; POMS_TMD = POMS Total Mood Disturbance

***Stage 3 Summary.*** The results of the above series of regression analyses indicate that the overall predictive ability of FAST to estimate performance impairments due to chronic sleep restriction is rather low. This conclusion is in stark contrast to the results reported by Chandler, Arnold, Phillips, Lojewski, and Horning (2010) for a study involving total sleep deprivation. Specifically, during total sleep deprivation FAST scores were able to predict nearly 14% of the variance in PVT lapses, but in the present study these scores accounted for only 1.1% of the variance. Moreover, even when constraining the data using different moderating factors such as personality facets, the predictive value of FAST never exceeded 10%. Since chronic sleep restriction is more common than total sleep deprivation in operational environments (Caldwell, Chandler, & Hartzler, 2012), the results of the above analyses suggest that FAST performance estimates should not be used as the sole means of determining readiness for duty.

The results of these regression analyses did reveal several factors which significantly improved the predictive value of the FAST. Across all analyses, the three best predictive factors during chronic sleep restriction appear to be an individual's scores on the POMS Fatigue/Inertia and Total Mood Disturbance scales, as well as their total lapse time on a flight simulator. This finding suggests that scores such as these, as well as PVT performance, may be valuable tools to shift managers and schedulers in determining whether personnel are ready for duty. Successful use of these types of algorithms could help to improve both mission efficacy and safety by identifying personnel who are best able to maintain performance under different operational conditions.

It is interesting that constraining the data prior to some of these analyses revealed major improvements to the original algorithm. For example, when the data were constrained by either participants' Gregariousness or Activity score on the NEO-PI-R, it was clear that the resulting algorithms were much stronger for participants ranked in the highest or lowest 25% of the facet. One possible explanation for this ties back to previous research which determined that these two Extraversion facets were related to fatigue susceptibility, with participants who scored higher on Extraversion traits also demonstrating greater susceptibility (Killgore et al., 2007). Although this effect was not significant in the present study (see Table 4 above), the trend indicated that the same was true for individuals undergoing chronic sleep restriction, with people who are outgoing and socially active appearing to be more fatigue susceptible whereas introverts were more fatigue resistant (see Figures 3 and 4 above). Thus it is possible that members of these respective groups may have had less inter-individual variability in response to sleep restriction, consequently making it easier for just a few significant factors to explain the majority of the variance. Similar results were obtained by constraining the data by the time of the testing session. Specifically, for the 0730 test session, almost 23% of the variance in the number of PVT lapses was accounted with just three factors, though the same was not true for the 1530 or 2330 testing sessions. This may be attributed to most or all participants succumbing to an early morning trough in the circadian cycle, which was further exacerbated by their growing sleep debt. Taken together, these results add further support to the notion that consideration of individual differences, as well as circadian factors, is crucial to effective scheduling practices.

## GENERAL DISCUSSION

### General Summary

Though total sleep deprivation is more commonly studied, chronic sleep restriction is widely regarded as the most common form of sleep loss experienced by modern society (Lindsay & Dyche, 2012; May & Kline, 1987). Despite countless studies identifying the many performance impairments which accompany either type of sleep loss, fatigue remains one of the most pervasive and potentially most devastating causes of human-factors related accidents and mishaps (Naval Safety Center, 2014). Thus, the present study was intended to examine the efficacy of several cognitive and physiologic measures for predicting performance impairments due to chronic sleep restriction. As a whole, the results of the present study support the idea that an effective fitness for duty screener could be developed by incorporating a variety of measures. Specifically, measures such as the Profile of Mood States (POMS) Fatigue-Inertia and Total Mood Disturbance scores, the total lapse time while on the flight simulator, as well as PMI Saccadic Velocity and Fatigue Avoidance Scheduling Tool (FAST) scores for predicted effectiveness, were sensitive not only to the group effects of fatigue but also to the individual differences in fatigue susceptibility. Conversely, many of the cognitive measures such as Flight Fit's Attention Shifting and Visual Scanning, as well as the dual *n*-back, indicated an improvement in participants' performance across the duration of the study, masking any fatigue-related effects.

### Comparing and Contrasting Chronic Sleep Restriction with Total Sleep Deprivation

The experiment reported herein is a follow-up to a total sleep deprivation (TSD) study completed by this laboratory (Chandler, Arnold, Phillips, Lojewski, & Horning, 2010). In particular, the TSD study was also designed to test the efficacy of several instruments as

potential fatigue detection tools and determine whether they were sensitive to individual differences. Results of that study revealed that only responses to the Stanford Sleepiness Scale, Flight Fit's Shifting Accuracy, and Saccadic Velocity as measured by the PMI FIT 2000, were sensitive to individual differences. Additionally, two reaction time measures (Flight Fit Raw RT and Divided Attention RT), as well as participants' performance on a flight simulator, also indicated impaired performance as a result of 25 hours of continuous wakefulness, though these measures were not sensitive to individual differences. Thus, although there are some similarities in participants' responses to the two types of sleep loss, such as Saccadic Velocity, there appear to be more differences. For example, participants' performance on four of the Flight Fit tasks deteriorated significantly during TSD, whereas performance improved significantly for most of the same measures during chronic sleep restriction (CSR).

On the surface, these conclusions might appear to be in contrast with the findings of Rupp, Wesensten, and Balkin (2012), which indicated that both TSD and CSR resulted in very similar performance impairments. That is, Rupp and colleagues exposed participants to both TSD conditions (continuous wakefulness for 63 hours) as well as CSR (restricted to 3-hrstime-in-bed for seven nights). The authors reported that not only did the response to sleep loss appear to be a stable trait across both types of sleep insults, but that performance decrements were similar for both conditions. For example, participants' performance on the Running Memory task, a measure of working memory similar to the dual *n*-back used in the present study, was similar for both sleep loss conditions.

An explanation for this difference in findings may be taken from a hypothesis put forth by Hursh and Van Dongen (2010) regarding the homeostatic set-point of sleep, suggesting that the similarity in performance decrements depends on the severity of the CSR. Specifically, Hursh and Van Dongen explained that CSR of 3 hours or less per night will result in continuous accumulation of performance impairment which is similar to, but slower than that of TSD. However, when sleep is reduced to 4 – 6 hours per night, the rate of decline will be much more gradual, eventually resulting in depressed but steady performance. The 4 hour mark which distinguishes the two levels of CSR has been termed the homeostatic "set-point". Taken together, the results of this research regarding TSD and CSR, as well as those obtained by Rupp and colleagues (2012) appear to lend credence to the set-point theory.

**Individual Differences in Assessing Fatigue**

As stated previously, fatigue due to sleep loss continues to be one of the foremost safety concerns in aviation, both for military and commercial operations. As part of a growing concern regarding fatigue, increasing interest has also focused on individual differences in fatigue susceptibility. These differences are static traits and may be related to a number of different factors, such as genetics, age, personality, and baseline neural activity. Because current guidelines are often based upon averages taken from groups, some individuals may actually be much more resistant to fatigue, or much more susceptible to fatigue. Failure to take these individual differences into consideration when establishing work schedules will likely lead to insufficient use of a given Warfighter's ability to function at acceptable levels with inadequate sleep, or place another at risk when his or her performance deteriorates quickly. Some of the measures used in this study may prove to be beneficial at not only assessing someone's present

level of fatigue, but also to predict how susceptible a particular individual may be hours or even days into the future.

**Fatigue Prediction Using Physiologic and Cognitive Measures**

As stated previously, the results of Stages 2 and 3 analyses revealed five measures which were sensitive to individual differences in fatigue susceptibility due to chronic sleep restriction and which may be used to predict future impairments. These measures included subjective (POMS Fatigue / Inertia and Total Mood Disturbance) and objective (total lapse time in a flight simulator and PMI FIT Saccadic Velocity) factors, as well as predictive effectiveness scores from the FAST. Taken together, these results suggest that the different types of factors should be used synergistically to develop a more accurate predictive algorithm which may better enable schedulers and commanders to determine whether a particular crew member is fit for duty.

Conversely, there were several measures which are highly sensitive to the performance impairments associated with total sleep deprivation but on which participants actually improved during chronic sleep restriction. These measures include any tasks that are likely to have performance effects, such as those involving memory or cognitive throughput. If the improvements had been confined to just the first few test sessions in the present study, it would have been easy to conclude that these conflicting results were simply due to practice effects which might have been prevented by including more training sessions. However, performance on several measures, such as the Flight Fit Visual Scanning Response Time (Figure 15) and Divided Attention Response Time (Figure 17), indicated that participants continued to improve even into the final day of testing. This suggests that even when obtaining sufficient sleep, participants would continue to improve their performance on these types of tests despite numerous administrations. Thus, the best measures for assessing and predicting fatigue due to chronic sleep restriction appear to be saccadic velocity, detailed subjective assessments such as the POMS, and measures of alertness and attention such as lapses on a flight simulator or the PVT.

**Future Directions**

The results of the study described herein provided a great deal of information not only about the manner in which restricting nightly time in bed to 4 hours influences cognitive and physiologic performance, but also how individuals vary in their degree of impairment. These individual differences may be exploited to improve existing scheduling practices by offering personnel better estimates of when their performance will be most impaired. However, before these types of predictive algorithms should be implemented in the field, further research is needed to determine which factors have the greatest predictive value at different levels of sleep loss. For example, as discussed above, impairments evident with fewer than 3 hours of sleep per night are highly similar to that observed under total sleep deprivation, whereas obtaining 4 or more hours of sleep leads to different impairments and even allows for some improvements in performance. Thus, subsequent studies should focus on identifying individual differences in fatigue susceptibility for participants receiving only 2 or 3 hours of sleep, as well as 5 or 6 hours of sleep, to determine which factors have the greatest predictive value for each level of sleep restriction. Further, the most successful fatigue detection tool would likely incorporate a wide variety of measures, including subjective and objective factors. Consequently, follow-up studies

should consider a wider variety of measures, such as more cognitive and physiologic assessments.

## REFERENCES


Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., . . . Wesensten, N. J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research, 13*, 219-227.

Banks, S. & Dinges, D. F. (2007). Behavioral and physiological consequences of sleep restriction. *Journal of Clinical Sleep Medicine, 3*, 519-528.

Banks, S. & Dinges, D. F. (2011). Chronic sleep deprivation. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Basic Edition* (5th ed., pp. 67-75). St. Louis: Saunders.

Belland, K. (2012, October). *Spatial disorientation* [PowerPoint slides]. Presented at the American Osteopathic College of Occupational and Preventative Medicine 2012 Annual Meeting, San Diego, CA. Retrieved from http://www.aocopm.org/assets/documents/OMED_2012_San_Diego/e-spatial%20disorentation%20dr%20belland.pdf

Bliese, P. D., Wesensten, N. J., & Balkin, T. J. (2006). Age and individual variability in performance during sleep restriction. *Journal of Sleep Research, 15*, 376-385.

Caldwell, J. A. (2001). The impact of fatigue in air medical and other types of operations: A review of fatigue facts and potential countermeasures. *Air Medical Journal, 20*, 25-32.

Caldwell, J. A. & Caldwell, J. L. (2005). Fatigue in military aviation: An overview of U.S. military-approved pharmacological countermeasures. *Aviation, Space, and Environmental Medicine, 76* (7, Supp.), 39-51.

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. *Aviation, Space, and Environmental Medicine, 80*, 29-59.

Caldwell, J. A., Mu, Q., Smith, J. K., Mishory, A., Caldwell, J. L., Peters, G., Brown, D. L., & George, M. S. (2005). Are individual differences in fatigue vulnerability related to baseline differences in cortical activation? *Behavioral Neuroscience, 119*, 694-707.

Caldwell, J. L., Chandler, J. F., & Hartzler, B. M. (2012). Battling fatigue in aviation: Recent advancements in research and practice. *Journal of Medical Science, 32*, 47-56.

Chandler, J. F., Arnold, R. D., Phillips, J. B., Lojewski, R. A., & Horning, D. S. (2010). *Preliminary validation of a readiness-to-fly assessment tool for use in Naval Aviation.* NAMRL Technical Report 10-22, Naval Aerospace Medical Research Laboratory, Pensacola FL.

Costa, P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Manual.* Odessa, FL: Psychological Assessment Resources.

Davenport, N. & Lee, J. (2007). Assessing how fatigue causes mishaps. *Approach, 52*, 6-9.

Dement (Eds.), *Principles and Practice of Sleep Medicine* Killgore, W. D. S., Grugle, N. I., Reichardt, R. M., Killgore, D. B., & Balkin, T. J. (2009). Executive functions and the ability to sustain vigilance during sleep loss. *Aviation, Space, and Environmental Medicine, 80*, 81-87.

Doty, R. L., McKeown, D. A., Lee, W. W., & Shaman, P. (1995). A study of the test-retest reliability of 10 olfactory tests. *Chemical Senses, 10*, 645-656.

Drury, D. A., Ferguson, S. A., & Thomas, M. J. W. (2012). Restricted sleep and negative affective states in commercial pilots during short haul operations. *Accident Analysis and Prevention, 455*, 80-84.

Gore, R. K., Webb, T. S., & Hermes, E. D. A. (2010). Fatigue and stimulant use in military fighter aircrew during combat operations. *Aviation, Space and Environmental Medicine, 81*, 719-727.

Greeley, H. P., Berg, J., Friets, E., Wilson, J., Greenough, G., Picone, J., Whitmore, J., & Nesthus, T. (2007). Fatigue estimation using voice analysis. *Behavioral Research Methods, 39*, 610-619.

Haack, M. & Mullington, J. M. (2005). Sustained sleep restriction reduces emotional and physical well-being. *Pain, 119*, 56-64.

Hartzler, B. M. (2014). Fatigue on the flight deck: The consequences of sleep loss and the benefits of napping. *Accident Analysis and Prevention, 32*, 309-318.

Hoddes, E., Zarcone, E., & Dement, W. (1972). Development and use of Stanford Sleepiness Scale (SSS). *Psychophysiology, 9*, 150.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T., J., . . . Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine, 7*, A44-A53.

Hursh, S. R. & Van Dongen, H. P. A. (2010). Fatigue and performance modeling. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine* (5th ed., pp. 745-752). Philadelphia: Elsevier Saunders.

Kendall, A. P., Kautz, M. A., Russo, M. B., & Killgore, W. D. S. (2006). Effects of sleep deprivation on lateral visual attention. *International Journal of Neuroscience, 116*, 1125-1138.

Killgore, W. D. S., Balkin, T. J., & Wesensten, N. J. (2006). Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research, 15*, 7-13.

Killgore, W. D. S., Killgore, D. B., Grugle, N. L., & Balking, T. J. (2010). Odor identification ability predicts executive function deficits following sleep deprivation. *International Journal of Neuroscience, 120*, 328-334.

Killgore, W. S., Richards, J. M., Killgore, D. B., Kamimori, G. H., & Balkin, T. J. (2007). The trait of introversion-extraversion predicts vulnerability to sleep deprivation. *Journal of Sleep Research, 16*, 354-363.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress, 3*, 129-141.

LeDuc, P. A., Riley, D., Hoffman, S. M., Brock, M. E., Norman, D., Johnson, P. A., Williamson, R., & Estrada, A. (1999). The effects of sleep deprivation on spatial disorientation. *USAARL Technical Report 2000-09.* DTIC AD Number: ADA 372369.

Lindsay, D. R. & Dyche, J. (2012). Sleep disturbance implications of modern military operations. *Journal of Human Performance in Extreme Environments, 10*, Article 2.

May, J. & Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *British Journal of Psychology, 78*, 443-455.

McNair, D.M., Lorr, M. & Droppleman, L.F. (1981). *Manual for the profile of mood states*. San Diego: Educational and Industrial Testing Service.

Naval Safety Center. (2014). Aeromedical factors involved in Naval Aviation class A flight mishaps from FY99-13. Unpublished briefing slide. Norfolk, VA.

Rogé, J., Pébayle, T., El Hannachi, S., & Muzet, A. (2003). Effect of sleep deprivation and driving duration on the useful visual field in younger and older subjects during simulator driving. *Vision Research, 43*, 1465-1472.

Rowland, L. M., Thomas, M. L., Thorne, D. R., Sing, H. C., Krichmar, J. L., Davis, H. Q., Balwinski, S. M., Peters, R. D., Kloeppel-Wagner, E., Redmond, D. P., Alicandri, E., & Belenky, G. (2005). Oculomotor responses during partial and total sleep deprivation. *Aviation, Space, and Environmental Medicine, 76* (Suppl. 7), C104-C113.

Rupp, T. L., Wesensten, N. J., Bliese, P. D., & Balkin, T. J. (2009). Banking sleep: Realization of benefits during subsequent sleep restriction and recovery. *Sleep, 32*, 311-321.

Rupp, T. L., Wesensten, N. J., & Balkin, T. J. (2012). Trait-like vulnerability to total and partial sleep loss. *Sleep, 35*, 1163-1172.

Sato, T., Kawada, T., Ogawa, M., Aoki, S., & Suzuki, S. (1993). Effects of some factors on sleep polygraphic parameters and subjective evaluations of sleep. *Environmental Research, 61*, 337-348.

Van Dongen, H. P. A., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep, 27*, 423-433.

Van Dongen, H. & Belenky, G. (2009). Individual differences in vulnerability to sleep loss in the work environment. *Industrial Health, 47*, 518-526.

Van Dongen, H. P. A. & Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinics in Sports Medicine, 24*, 237-249.

Van Dongen, H. P. A., Caldwell, J. A., & Caldwell, J. L. (2006). Investigating systematic individual differences in sleep-deprived performance on a high-fidelity flight simulator. *Behavior Research Methods, 38*, 333-343.

Van Dongen, H. P. A., Vitellaro, K. M., & Dinges, D. F. (2005). Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep, 28*, 479-496.

Venkatraman, V., Chuah, Y. M. L., Huettel, S. A., & Chee, M. W. L. (2007). Sleep deprivation elevates expectation of gains and attenuates response to losses following risky decisions. *Sleep, 30*, 603-609.